

RECONNAISSANCE DE LOCUTEURS
EN SCIENCES FORENSIQUES :

L'APPORT D'UNE APPROCHE
AUTOMATIQUE

Thèse de doctorat

Présentée à

l'Institut de Police Scientifique et de Criminologie
de l'Université de Lausanne

par

Didier Meuwly

Licencié en sciences forensiques
de l'Université de Lausanne

**Institut de police scientifique
et de criminologie**

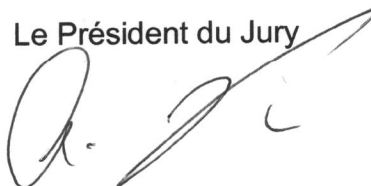
UNIL - Bâtiment de Chimie
CH - 1015 LAUSANNE-DORIGNY
Tél. 021/692 46 00
Fax 021/692 46 05

IMPRIMATUR

A l'issue de la soutenance de thèse, le Jury autorise l'impression de la thèse de Monsieur Didier MEUWLY, candidat au doctorat en sciences forensiques, intitulée

**«Reconnaissance automatique de locuteurs en sciences forensiques,
l'apport d'une approche automatique»**

Le Président du Jury



Professeur André KUHN

Lausanne, le 22 mai 2000

À mes parents

À Nicole

*Les sons émis par la voix sont les symboles des états de l'âme,
et les mots écrits, les symboles des mots émis par la voix.
Et de même que l'écriture n'est pas la même chez tous les hommes,
les mots parlés ne sont pas non plus les mêmes,
bien que les états de l'âme dont ces expressions
sont les signes immédiats soient identiques chez tous,
comme sont identiques aussi les choses dont ces états sont les images.*

De interpretatione, 1, 16, a 5 - 10 ; ARISTOTE (384 - 322 av. J.-C.)

REMERCIEMENTS

Ce travail de thèse a été réalisé à l'Institut de Police Scientifique et Criminologie (IPSC) de la Faculté de Droit de l'Université de Lausanne. La direction de la thèse a été assurée conjointement par le Docteur Andrzej Drygajlo, responsable du traitement de la parole au Laboratoire de Traitement des Signaux (LTS) de l'École Polytechnique Fédérale de Lausanne et le Professeur Pierre Margot, directeur de l'Institut de Police Scientifique et de Criminologie.

Le jury de thèse était composé de Monsieur le Professeur André Kuhn, Professeur associé à la Faculté de Droit de l'Université de Lausanne, président du jury, de Monsieur le Professeur Pierre Margot, Directeur de l'Institut de Police Scientifique et de Criminologie, rapporteur, de Monsieur le Docteur Frédéric Bimbot, chargé de recherche au CNRS, expert, et de Monsieur le Docteur Ton Broeders, responsable du *Department of Writing and Speech* du laboratoire national de sciences forensiques des Pays-Bas, expert.

Je tiens à exprimer ici ma vive gratitude et mes sincères remerciements à toutes les personnes qui m'ont apporté leur amitié, leur connaissance et leur aide tout au long de ce travail, et en particulier :

À mon directeur de thèse, Monsieur le Docteur Andrzej Drygajlo, pour m'avoir accepté sous son aile et m'avoir fait découvrir avec patience et gentillesse le monde et quelques-uns des secrets du traitement du signal de la parole ;

À Monsieur le Professeur Pierre Margot, pour toutes les possibilités qu'il m'a offertes de découvrir et partager sa passion des sciences forensiques ;

À Monsieur le Docteur Frédéric Bimbot, pour l'intérêt qu'il porte à l'application forensique de la reconnaissance de locuteurs et la rigueur scientifique qui anime sa réflexion ;

À Monsieur le Docteur Ton Broeders, pour m'avoir guidé sur le difficile chemin de l'expertise de reconnaissance de locuteurs par son ouverture d'esprit et sa grande expérience scientifique ;

À Monsieur le Professeur Christophe Champod, pour sa disponibilité de tous les instants et sa propension naturelle à partager ses immenses connaissances dans le domaine de l'interprétation de la preuve scientifique ;

À Monsieur Mounir El Maliki, doctorant au laboratoire de traitement des signaux de l'École Polytechnique Fédérale de Lausanne, pour avoir su déchiffrer mes explications et les retranscrire en lignes de code efficaces, avec un humour et une bonne humeur à toute épreuve ;

À Monsieur Philippe Renevey, doctorant au laboratoire de traitement des signaux de l'École Polytechnique Fédérale de Lausanne, pour sa dextérité dans l'art du pilotage des stations UNIX ;

À Monsieur Robert van Kommer, responsable du laboratoire *R & D Digital Signal Processing* de *Swisscom*[®], pour m'avoir mis gracieusement à disposition la base de données « Polyphone Suisse Romande » ;

À Monsieur Philippe Schucany, chef du Service d'identification de la Police Cantonale de Neuchâtel, pour avoir ouvert les portes de son service à mes expériences et m'avoir accordé sa confiance dans plusieurs affaires criminelles ;

À Monsieur le Professeur François Grosjean, Professeur de phonétique à l'Université de Neuchâtel, pour m'avoir accueilli dans son cours de phonétique acoustique ;

À Monsieur le Professeur Eric Keller, Professeur d'informatique et méthodes mathématiques à l'Université de Lausanne, pour m'avoir gracieusement mis à disposition son logiciel « *Signalize*TM » ;

À Monsieur Jean-Pierre Rosset, responsable de la partie audio du centre audiovisuel de l'Université de Lausanne, pour ses conseils et son aide précieuse dans la numérisation d'enregistrements sonores ;

À Monsieur le Docteur Tony Cantu, responsable de la recherche scientifique de la *Forensic Services Division of the United States Secret Service*, qui, avec enthousiasme, m'a assuré de précieux contacts aux États-Unis et procuré la littérature nord-américaine pertinente ;

À Monsieur Steve Lewis, du *Home Office* de Grande-Bretagne, que je n'ai pas la chance de connaître, mais dont la réflexion philosophique sur la reconnaissance de locuteurs est à la base de ce travail ;

À toutes les personnes qui œuvrent dans les bibliothèques des Universités et des Écoles Polytechniques Fédérales de Suisse et qui, par leur travail et leur disponibilité, contribuent à mettre en valeur et à rendre accessible les trésors insoupçonnés qui s'y cachent ;

Aux personnes qui, en prêtant leur voix et leur temps, ont contribué à la constitution de la base de données « Polyphone IPSC ». Qu'elles soient ici remerciées de leur persévérance et de leur disponibilité. Ce sont :

Mesdames Caroline et Ruth Behr, Anne Brunelle, Dominique et Fabienne Emonet, Barbara et Caroline Lauber, Eliane et Geneviève Massonnet, Monique Mermilliod, Nicole et Rose-Marie Meuwly, Agatina et Fortinata Santangelo, Martine Tristan-Udriot, Valérie Tristan-Rochaix et Messieurs Alexandre et Maurice Boin, Jean-François et Marc Chevalley, Marcel et Raphaël Coquoz, Marc et Robert Demierre, Alexandre et Marc Girod, Jacques et Pierre Mathyer, Bernard Meuwly ainsi que Jean-Michel et Pierre-Louis Rochaix.

À Madame Suzanne Dieterle et Madame le Docteur Geneviève Massonnet ainsi que Monsieur Bernard Meuwly, pour leur relecture attentive de ce manuscrit et leurs suggestions pertinentes.

À mes parents, pour leur soutien inconditionnel dans tout ce que j'ai entrepris et que j'ai pu réussir grâce à eux.

À Nicole, ma petite sœur, pour sa soif et sa joie de vivre, un rayon de soleil dans le monde des écrans cathodiques.

À Monsieur l'abbé Georges Rukundo, mon ami, pour l'exemple de courage qu'il m'a donné, lui qui a survécu à la justice inique et aux prisons de son pays.

À Tacha, ma collègue de bureau et amie, qui durant six ans a constaté avec moi que l'arbre de la Connaissance pousse très lentement.

Je tiens également à remercier tous mes collègues et amis de l'Institut de Police Scientifique et de Criminologie:

Frédéric Anglada, Alexandre Anthonioz, Monica Bonfanti, Julien Cartier, Michèle Claude, Raphaël Coquoz, Olivier Delémont, Eric Dupasquier, Eric Dürst, Pierre Esseiva, Françoise Fridez, Alain Gallusser, Aïta Khanmy-Vital, Eric Lock, Jean-Claude Martin, William Mazzella, Florence Monard-Sermier, Cédric Neumann, Joëlle Papilloud, Christophe Reymond, Olivier Ribaux, Eric Sapin, Franco Taroni et Christian Zingg.

AVANT-PROPOS

Le titre de cette recherche mérite tout d'abord une explication. Les sciences forensiques constituent l'ensemble des principes scientifiques et des méthodes techniques appliquées à l'investigation criminelle, pour prouver l'existence d'un crime et aider la justice à déterminer l'identité de l'auteur et son mode opératoire. La reconnaissance automatique de locuteurs s'intéresse aux processus de décision informatisés, qui utilisent quelques caractéristiques du signal de parole pour déterminer si une personne particulière est l'auteur d'un énoncé donné.

La reconnaissance automatique de locuteurs est relativement méconnue du grand public, car personne n'a l'expérience de son utilisation dans la vie de tous les jours, dans un système de contrôle d'accès à des services bancaires par exemple. Ce constat en demi-teinte dénote l'aspect encore expérimental de cette technologie, malgré le nombre important de recherches entreprises dans ce domaine depuis bientôt quarante ans.

Pourtant, les acteurs du monde judiciaire se prononcent régulièrement en faveur de l'utilisation de la reconnaissance de locuteurs dans le cadre de l'investigation criminelle et de l'expertise judiciaire. Plusieurs raisons peuvent expliquer cette prise de position. La principale est certainement la sous-estimation de la difficulté de la procédure de reconnaissance de locuteurs par la personne inexperte. Forte de son expérience dans l'identification de ses proches par la voix, elle est persuadée que cette constatation demeure valide en toute situation. Une autre explication provient certainement du cinéma, de la télévision et de la littérature, qui nourrissent l'idée qu'il existe des techniques scientifiques validées et fiables permettant la reconnaissance de locuteurs dans n'importe quelle circonstance.

Cette réalité ambivalente est à l'origine de la présente étude. Elle s'adresse avant tout au criminaliste, censé connaître l'ensemble des méthodes scientifiques d'identification et conseiller de manière pertinente les acteurs du monde judiciaire.

Cette recherche tente de fournir une vue d'ensemble des méthodes de reconnaissance de locuteurs utilisées aujourd'hui dans le domaine forensique et d'en saisir les enjeux et les limites. Pour y parvenir, le présent ouvrage est structuré en quatre parties. Après une approche théorique, il propose une analyse des procédures utilisées en sciences forensiques, se poursuit par une recherche expérimentale destinée à évaluer l'apport d'une approche automatique dans ce domaine et se termine par une discussion générale et une conclusion en forme de synthèse.

Nous souhaitons à la lectrice et au lecteur de trouver autant de plaisir à la lecture de cet ouvrage que nous en avons eu à la réalisation de ce projet.

RESUME

Cette recherche tente de fournir une vue d'ensemble des méthodes de reconnaissance de locuteurs utilisées aujourd'hui dans le domaine forensique et d'en saisir les enjeux et les limites. Pour y parvenir, le présent ouvrage est structuré en quatre parties.

L'approche théorique rappelle les approches classiques, inductive et déductive, utilisées pour l'identification en sciences forensiques et explore la voix en tant qu'indice matériel. Elle propose une méthodologie nouvelle, basée sur le théorème de Bayes, comme canevas d'interprétation pour la reconnaissance de locuteurs en sciences forensiques. Cette méthodologie permet l'évaluation de la vraisemblance de l'indice matériel dans deux hypothèses alternatives : premièrement dans l'hypothèse que la source de l'indice est le locuteur suspecté et deuxièmement dans une hypothèse alternative dans laquelle le locuteur suspecté n'est pas la source de l'indice.

La recherche bibliographique définit l'état de l'art et analyse les trois approches utilisées pour la reconnaissance de locuteurs en sciences forensiques: l'approche auditive, l'approche spectrographique et l'approche automatique.

La recherche expérimentale décrit le développement d'un système automatique de reconnaissance de locuteurs basée sur la méthode de modélisation par mélange de fonctions de densité gaussiennes (GMM-Gaussian Mixture Models) et le développement d'une approche continue du calcul des rapports de vraisemblance, notamment grâce par l'estimation de densité de noyaux (KDE-Kernel Density Estimation). Le système ainsi développé est ensuite testé dans diverses conditions typiquement rencontrées en sciences forensiques comme : l'influence de la qualité et la quantité des données, l'influence d'un déguisement de la voix, l'influence de la ligne et du téléphone, l'influence du bruit de fond, l'influence du système d'enregistrement et l'influence de voix auditivement proches.

Le bilan de la recherche et la question de l'utilisation dans la réalité de la reconnaissance de locuteurs en sciences forensiques sont développées dans la discussion générale et la conclusion, rédigées en forme de synthèse.

ZUSAMMENFASSUNG

Diese Forschung versucht eine Gesamtübersicht der heute im Gebiete der Forensik angewandten Methoden zur Sprechererkennung zugeben sowie die damit verbundenen Risiken und Limiten zu erfassen. Dazu ist die vorliegende Arbeit in vier Abschnitte gegliedert.

Im theoretischen Teil werden die in den forensischen Wissenschaften angewandten klassischen, induktiven und deduktiven Identifikationsmethoden beschrieben. Die Stimme als Materialbeweis wird erforscht und eine neue auf dem Theorem von Bayes basierende Methodologie als Interpretationsanleitung zur forensischen Sprechererkennung vorgeschlagen. Diese Methodologie erlaubt die Wahrscheinlichkeit des Materialindizes in bezug auf zwei Alternativhypothesen abzuschätzen, nämlich: einerseits die Hypothese, dass die Quelle des Indizes der Sprecher ist, und andererseits, die alternative Hypothese, dass der Sprecher nicht die Quelle des Indizes ist.

Der Literaturüberblick informiert über den Stand der Forschung und analysiert die drei für die forensische Sprechererkennung verwendeten Methoden, nämlich die auditive Methode, die spektrographische Methode und die automatische Methode.

Der Experimentaltteil beschreibt die Entwicklung eines automatischen Sprechererkennungssystems, welches auf dem GMM-Gaussian Mixture Model- basiert, sowie einer Methode zur einer fortlaufenden Berechnung des Wahrscheinlichkeitsverhältnisses, im besonderen durch die KDE-Kernel Density Estimation. Das so entwickelte System wurde unter verschiedenen, spezifisch in der Forensik angetroffenen Bedingung getestet, wie der Einfluss der Qualität und der Quantität der Daten, der Einfluss der Stimmverstellung, der Einfluss der Leitung und des Telephons, der Einfluss des Hintergrundrauschens, der Einfluss des Aufnahmesystems und der Einfluss von auditiv nahen Stimmen.

Die Bilanz der Forschung und die Frage nach der realen Anwendung in der forensischen Sprechererkennung werden in der allgemeinen Diskussion und der in Form einer Synthese formulierten Schlussfolgerung behandelt.

SUMMARY

This study attempts to present a synthesis on the methods currently used in forensic science for identifying speakers, to define the issues involved, and to assess practical limitations. To this end we have divided the present work into four parts.

A theoretical discussion summarises established methods, both inductive and deductive, practised in forensic science for purposes of identification, examining the voice as a material trace. A new methodology is proposed, based on Bayes's theorem, as a framework for interpretation in speaker recognition for forensic science. This methodology permits an assessment of the probabilities for trace material following two alternative hypotheses: we first consider the source of the trace as the suspected speaker; in the second hypothesis the suspected speaker is not the source of the trace.

Our second part is a bibliography outlining the state of the discipline and analyses the three methods used for recognition of speakers in forensic applications: the aural approach, the spectrographic approach and automatic speaker recognition.

Our experimental section describes the development of an automatic system for speaker recognition and establishes a model following a mixture of Gaussian density functions (Gaussian Mixture models-GMM). We use a continuous approach in calculating probability ratios, particularly as regards kernel density estimation (KDE). The system proposed is then tested under several conditions typically encountered in forensics, such as: influence of quality and quantity of data, influence of attempts to disguise the voice, influence of the telephone line and handset, background noise, recording system, and consideration of voices which appear to share aural proximity.

The results of our investigation and the question of practical use in forensic science in recognising speakers are set forth in our general discussion and conclusion, where we attempt a synthesis.

SOMMAIRE

PARTIE 1 : APPROCHE THEORIQUE

- I. Introduction
- II. La voix comme indice matériel
- III. Méthodologie

PARTIE 2 : RECHERCHE BIBLIOGRAPHIQUE

- IV. Approche auditive
- V. Approche spectrographique
- VI. Approche automatique

PARTIE 3 : RECHERCHE EXPERIMENTALE

- VII. Développement d'un système automatique de reconnaissance de locuteurs
- VIII. Évaluation du système

PARTIE 4 : SYNTHESE

- IX. Discussion générale
- X. Conclusion

Annexes

Bibliographie

TABLE DES MATIERES

Sommaire	I
Table des matières	III
PARTIE 1 : APPROCHE THEORIQUE	1
I. Introduction	3
1.1. La notion d'identité en sciences forensiques	3
1.1.1. Définitions	3
1.1.2. Concepts et raisonnement	3
1.2. La voix comme caractère d'identité	4
1.2.1. La voix humaine	4
1.2.2. La voix comme moyen d'identification	5
1.3. Le rôle des probabilités dans l'identification	5
1.3.1. Définitions	5
1.3.2. Limites de l'approche subjective	6
1.3.3. Limites de l'approche statistique	7
1.4. Hypothèse de la recherche	7
1.5. Objectifs de la recherche	7
1.6. Contributions majeures	8
1.7. Organisation de la recherche	9
II. La voix comme indice matériel	11
2.1. Introduction	11
2.2. Cadre légal	12
2.2.1. Conditions de recevabilité d'un enregistrement téléphonique en Suisse	12
2.2.2. La procédure d'écoute téléphonique en Suisse	13
2.3. Collecte de l'indice matériel	14
2.3.1. Description et représentation du signal de parole	15
2.3.2. Mesure de la qualité de la parole	16
2.3.3. Influence du système de codage numérique de l'information	16
2.3.4. Influence de la prise de son	20
2.3.5. Influence du canal de transmission	20

2.3.6. Influence du système d'enregistrement	22
2.3.7. Influence du type d'investigation	23
2.3.8. Influence du locuteur	24
2.4. Conclusion	27
III. Méthodologie	29
3.1. Introduction	29
3.2. Rôle de l'expert ou du scientifique	29
3.2.1. Le refus de témoigner	29
3.2.2. Le maximalisme	29
3.2.3. La présentation et l'évaluation de l'état de l'art	30
3.2.4. Choix d'une approche méthodologique	30
3.3. Exigences légales en matière de preuve scientifique	30
3.3.1. En droit suisse	30
3.3.2. En droit nord-américain	31
3.3.3. Choix d'une démarche	32
3.4. Méthodes de reconnaissance de locuteurs	32
3.4.1. Définitions	32
3.4.2. Procédure	33
3.4.3. Classification des méthodes de reconnaissance	34
3.4.4. Choix d'une méthode	36
3.5. Inférence de l'identité d'un locuteur	37
3.5.1. Discrimination	37
3.5.2. Classification	39
3.5.3. Quantification des taux d'erreur de type I et de type II	42
3.5.4. Évaluation de rapports de vraisemblance	44
3.5.5. Choix d'un processus d'inférence de l'identité	47
3.6. Évaluation d'une méthode de reconnaissance automatique de locuteurs	50
3.6.1. Établissement de modèles théoriques	50
3.6.2. Comparaison de modèles théoriques	50
3.6.3. Évaluation empirique	50
3.6.4. Choix d'une méthode d'évaluation	51
3.7. Conclusion	52

PARTIE 2 : RECHERCHE BIBLIOGRAPHIQUE	53
IV. Approche auditive	55
4.1. La perception de la voix et de la parole	55
4.1.1. Principes de la perception	55
4.1.2. Le processus de discrimination et d'identification de locuteurs	55
4.2. Les méthodes de reconnaissance auditive	56
4.3. Procédure de reconnaissance par des profanes	56
4.3.1. Approche descriptive	57
4.3.2. Limites de l'approche descriptive	60
4.3.3. Approche expérimentale	61
4.3.4. Limites de la procédure de reconnaissance par des profanes	73
4.4. Procédure de reconnaissance par des experts	74
4.4.1. L'approche auditive perceptive	74
4.4.2. L'approche phonétique acoustique	77
4.4.3. Limites des approches auditive perceptive et phonétique acoustique	83
V. Approche spectrographique	85
5.1. Le spectrographe sonore	85
5.1.1. La technologie	85
5.1.2. L'application à la reconnaissance de locuteurs	85
5.2. L'application forensique	86
5.2.1. La méthode de KERSTA	86
5.2.2. Tentative de validation de la méthode de KERSTA : l'étude de TOSI	91
5.2.3. Recevabilité de la méthode spectrographique	100
5.3. Rapport du Conseil National des Sciences	103
5.3.1. Position du rapport sur les différents éléments de controverse	103
5.3.2. Conclusion du rapport du Conseil National des Sciences	105
5.4. Après le rapport du Conseil National des Sciences	106
5.4.1. La dissolution de l'IAVI	106
5.4.2. L'étude du FBI	106
5.4.3. Les standards de l'IAI	107
5.4.4. L'arrêt Daubert	108
5.5. La méthode spectrographique dans le reste du monde	110

5.6. Conclusion	111
VI. Approche automatique	113
6.1. Introduction	113
6.1.1. Définition	113
6.1.2. Historique	113
6.2. Analyse du signal de parole	114
6.2.1. Principes	114
6.2.2. Approches primaires	115
6.2.3. Approches actuelles	119
6.2.4. Conclusion	125
6.3. Mesure de similarité	126
6.3.1. Approches primaires	126
6.3.2. Approches actuelles	129
6.4. Systèmes automatiques développés en sciences forensiques	135
6.4.1. Semi-Automatic Speaker Identification System (SASIS) - USA (1971 - 1975)	135
6.4.2. Automatic Recognition Of Speakers (AUROS) – Allemagne (1977)	136
6.4.3. Computer Assisted Voice Identification System (CAVIS) - USA (1985 - 1989)	137
6.4.4. Semi-AUtomatic Speaker Identification system (SAUSI) - USA (1976-1998)	138
6.4.5. IDentification Method (IDEM) – Italie (dès 1991)	139
6.4.6. REconnaissance Vocale Assistée par Ordinateur (REVAO) – France (1988 – 1993)	139
6.4.7. Approches récentes	141
6.5. Conclusion	143
PARTIE 3 : RECHERCHE EXPERIMENTALE	145
VII. Développement d'un système automatique de reconnaissance de locuteurs	147
7.1. Introduction	147
7.2. Le système de reconnaissance de locuteurs	147
7.2.1. Définition générale du système	147
7.2.2. Architecture du système	148
7.2.3. Prétraitement du signal	149
7.3. Méthode de calcul du rapport de vraisemblance	152
7.3.1. Production des données	152
7.3.2. Distribution des données	152
7.3.3. Estimation de la distribution par <i>kernel density estimation</i>	153

7.3.4. Estimation des fonctions de densité de probabilité	155
7.3.5. Calcul du rapport de vraisemblance de l'élément de preuve E	156
7.4. Expériences	158
7.4.1. Principe	158
7.4.2. Présentation des résultats	159
7.4.3. Exemple	159
7.5. Conclusion	160
VIII. Évaluation du système	161
8.1. Introduction	161
8.2. Enregistrement et sélection de bases de données	161
8.2.1. Détermination de la langue parlée	161
8.2.2. Estimation de la variabilité intralocuteur	161
8.2.3. Estimation de la variabilité interlocuteur	164
8.2.4. Constitution d'enregistrements de test	166
8.3. Procédure d'évaluation du système	166
8.4. Limites théoriques du système	167
8.4.1. Évaluation sur la base de données " Polyphone Suisse Romande "	167
8.4.2. Évaluation sur la base de données " Polyphone IPSC "	168
8.4.3. Discussion des résultats	168
8.5. Évaluation de l'influence du temps séparant l'enregistrement de l'indice et celui du modèle	169
8.5.1. Procédure	169
8.5.2. Résultats	170
8.5.3. Discussion des résultats	171
8.6. Évaluation de l'influence de la qualité et de la quantité de données	171
8.6.1. Influence du type d'élocution lors de l'enregistrement des modèles	171
8.6.2. Influence de la quantité de parole dans les enregistrements de comparaison	173
8.6.3. Influence du type d'élocution dans les enregistrements de comparaison	175
8.7. Évaluation de l'influence d'un déguisement de la voix	177
8.7.1. Déguisement de la voix dans les enregistrements de comparaison	177
8.7.2. Déguisement de la voix dans les enregistrements de test	179
8.8. Évaluation de l'influence du réseau, de la ligne et du téléphone	182
8.8.1. Influence du téléphone et de la ligne téléphonique utilisés pour l'enregistrement des modèles	182
8.8.2. Influence du téléphone et de la ligne téléphonique utilisés pour les enregistrements de test	183

8.8.3. Influence du réseau utilisé pour l'enregistrement des modèles	184
8.8.4. Influence du réseau utilisé pour la production des enregistrements de test	187
8.9. Évaluation de l'influence du bruit de fond	188
8.9.1. Procédure	188
8.9.2. Résultats	189
8.9.3. Discussion des résultats	190
8.10. Évaluation de l'influence du système d'enregistrement des indices	191
8.10.1. Procédure	191
8.10.2. Résultats	191
8.10.3. Discussion des résultats	192
8.11. Évaluation de l'influence de voix auditivement proches	192
8.11.1. Influence du téléphone et de la ligne téléphonique	192
8.11.2. Influence du réseau téléphonique	194
8.11.3. Influence d'un déguisement de la voix	195
8.11.4. Discussion sur les voix auditivement proches	197
PARTIE 4 : SYNTHESE	199
IX. Discussion générale	201
9.1. Introduction	201
9.2. Bilan de la recherche	201
9.2.1. Réflexion sur la démarche	201
9.2.2. Réflexion sur les méthodes	203
9.2.3. Situation de l'approche spectrographique	207
9.2.4. Réflexion sur les résultats	208
9.2.5. Voies de recherche	209
9.3. Utilisation dans la réalité de l'approche automatique développée	210
9.3.1. Aspects méthodologiques	210
9.3.2. Aspects techniques	211
9.3.3. Aspects juridiques	211
9.3.4. Aspects d'organisation	212
X. Conclusion	213

Annexes	215
Annexe I. Extraits de la Constitution fédérale de la Confédération suisse (RS 101)	217
Chapitre premier: Dispositions générales	217
Titre 2: Droits fondamentaux, citoyenneté et buts sociaux	217
Chapitre premier: Droits fondamentaux	217
Annexe II. Extraits du Code pénal suisse (RS 311.0)	219
Livre premier: Dispositions générales	219
Première partie: Des crimes et des délits	219
Livre deuxième: Dispositions spéciales	220
Titre troisième: Infractions contre l'honneur et contre le domaine secret ou le domaine privé	220
Annexe III. Ordonnance sur le service de surveillance de la correspondance postale et des télécommunications (RS 780.11)	221
Section 1: Organisation	221
Section 2: Surveillance de la correspondance postale	221
Section 3: Surveillance des télécommunications	222
Section 4: Renseignements sur les raccordements	223
Section 5: Dispositions communes	224
Section 6: Dispositions finales	225
Annexe IV. Extraits des <i>Federal Rules of Evidence</i>	227
<i>Article I: General Provisions</i>	227
<i>Article VII: Opinions and Expert Testimony</i>	227
<i>Article IX: Authentication and Identification</i>	228
Annexe V. Code de procédure de l'International Association for Forensic Phonetics (IAFP)	231
Code de procédure	231
Annexe VI. Base de données "Polyphone IPSC"	233
A.VI.1. Date des sessions d'enregistrement	233
A.VI.2. Type de téléphone utilisé	235
A.VI.3. Composition des enregistrements	238
Bibliographie	247
Bibliographie	249

PARTIE 1

APPROCHE THEORIQUE

I. INTRODUCTION

1.1. La notion d'identité en sciences forensiques

1.1.1. Définitions

« L'identité est un concept humain qui découle directement de l'expérience du monde physique. Les objets, spécifiques ou génériques, sont groupés ou organisés dans la mémoire en fonction d'expériences passées de la perception de leurs caractéristiques intrinsèques. Comme il n'est possible d'appréhender le monde que par les sens, c'est un processus inductif qui gouverne les identités ainsi établies » [LEWIS, 1984].

« En police scientifique et en droit, l'identité est l'ensemble des caractères par lesquels un homme définit sa personnalité propre et se distingue de tout autre. Dans ce dernier ordre d'idées, établir l'identité d'un individu, est l'opération policière ou médico-légale appelée identification. Un homme peut être semblable à plusieurs autres, ou à un autre, au point d'amener des erreurs ; il n'est jamais identique qu'à un seul, à lui-même. C'est à discriminer avec soin les éléments de ressemblance des éléments d'identité que consiste le problème de l'identification » [LOCARD, 1909].

1.1.2. Concepts et raisonnement

1.1.2.1. L'identité

Une raison fondamentale de l'organisation de la mémoire est de permettre d'établir un lien entre expérience présente et expériences passées [LEWIS, 1984].

« Il n'y a pas d'observation naïve : tout ce que nous observons autour de nous est structuré par les expériences que nous avons faites, c'est-à-dire par les théories qui se sont confirmées jusqu'ici » [POPPER, 1988].

Une autre raison est la tentative de prédire ou d'inférer la connaissance associée à des expériences personnellement non tentées, soit dans le passé, soit dans le futur. Ces deux buts doivent être clairement distingués pour prévenir toute confusion dans la compréhension de la preuve d'identité [LEWIS, 1984]. Appréhender le présent sur la base d'expériences passées fait appel au raisonnement déductif. Comme ces déductions sont basées sur des identités établies de façon inductive, elles impliquent que les identités établies dans le passé sont correctes jusqu'à ce qu'une expérience contradictoire ne révèle l'inexactitude de ce raisonnement. Par conséquent les identités, et delà les relations établies sur la base de ces identités, peuvent évoluer au cours des expériences. Prédire le futur ou inférer la connaissance associée à des expériences personnellement non tentées, comme dans la situation forensique, s'appuie sur le raisonnement inductif et déductif.

Durant une procédure d'identification, il est essentiel de distinguer la part de la décision qui s'appuie sur le raisonnement inductif, de celle qui repose sur le raisonnement déductif. La contribution de la science se trouve dans le domaine du raisonnement déductif, mais les résultats

obtenus scientifiquement influencent souvent de manière prépondérante les conclusions du raisonnement inductif et déterminent fréquemment plusieurs des conditions nécessaires à leur obtention. Le raisonnement inductif est si intrinsèquement associé au processus d'identification que son rôle devrait être mentionné dans chaque procédure d'identification [LEWIS, 1984].

1.1.2.2. L'identification

En sciences forensiques, le processus d'identification vise à l'individualisation [KWAN, 1977 ; TUTHILL, 1994]. Identifier une personne ou un objet signifie qu'il est possible de les distinguer de toutes les autres personnes ou de tous les autres objets de la Terre. Ce processus peut être vu comme un processus de réduction d'une population initiale jusqu'à l'unité. La taille de la population initiale dépend des circonstances, mais elle comprend au maximum la population des êtres humains de la Terre ou de la catégorie d'objets considérés.

Le facteur de réduction est défini par la spécificité ou la rareté des caractéristiques concordantes observées entre un indice matériel, comme l'enregistrement d'un message anonyme, et un enregistrement de comparaison, par exemple l'enregistrement de la voix d'une personne mise en cause. La conclusion de l'identification est une opinion, l'expression d'une probabilité, subjective ou objective, indiquant que la chance d'observer sur la Terre une personne ou un objet présentant des caractéristiques concordantes tend vers zéro [CHAMPOD ET MEUWLY, 1998].

1.2. La voix comme caractère d'identité

1.2.1. La voix humaine

La production de la parole trouve sa source dans l'activité respiratoire, dont elle est dépendante, puisque l'appareil de la phonation ne possède pas d'individualité anatomique. Elle est le résultat de deux fonctions mécaniques de base : la phonation et l'articulation. La phonation consiste en la production d'un phénomène acoustique. L'articulation inclut la modulation de ce phénomène acoustique par les articulateurs, principalement les lèvres, la langue et le palais, ainsi que sa modulation, par les cavités supraglottiques, orales et/ou nasales [GUYTON, 1984]. L'énergie expiratoire est utilisée pour produire des bruits et/ou mettre en mouvement les cordes vocales, qui génèrent les sons voisés. Les sons de la parole peuvent être caractérisés dans les domaines temporel, spectral et spectro-temporel ; les unités segmentales de la parole, les phonèmes, se divisent en consonnes, semi-consonnes et en voyelles. La parole est l'un des premiers moyens de communication entre les êtres humains ; ce comportement est régi par un code, le langage.

L'étendue spectrale de la voix humaine est comprise entre 80 et 8000 Hz et la puissance sonore de la parole normale se situe de 60 à 70 dB. La fréquence fondamentale moyenne de vibration des cordes vocales (F_0) est comprise entre 180 et 300 Hz chez les femmes, entre 300 et 600 Hz chez les enfants et entre 90 et 140 Hz chez les hommes.

1.2.2. La voix comme moyen d'identification

La voix est une caractéristique biométrique comme le sont l'empreinte digitale, le réseau vasculaire rétinien ou l'information génétique. En tant que telle, elle bénéficie d'un *a priori* de très grande fiabilité, voire même d'infailibilité en terme d'identification [DODDINGTON, 1985]. Ce « principe d'individualité », particulièrement associé aux mesures biométriques, est souvent invoqué en sciences forensiques [ROBERTSON ET VIGNAUX, 1995] et parfois appliqué à la voix humaine [KÜNZEL, 1987 ; KLEVANS ET RODMAN, 1997]. Il n'est pourtant ni justifiable *a priori*, ni démontrable d'un point de vue théorique [KWAN, 1977].

Ce point de vue s'explique d'une part par le fait que, dans la plupart des procédures d'identification, l'identité dont il est question est basée sur des caractéristiques non statistiques et essentiellement figées. D'autre part, le processus inductif impliqué dans la comparaison de ces caractéristiques d'identité n'est en général ni conscient ni considéré de façon critique. Toutefois, il forme la base de l'observation scientifique d'une information comme la probabilité statistique d'occurrence d'une caractéristique d'identité dans une population [LEWIS, 1984].

La problématique de l'identification par la voix en est une parfaite illustration. Aucune caractéristique spécifique distinguant deux voix n'a été mise en évidence jusqu'à présent. Cet échec indique soit qu'effectivement le « principe d'individualité » n'est pas fondé, soit qu'il ne s'applique pas à la voix, soit que la limitation des méthodes de détection mises en œuvre jusqu'à présent n'a pas permis cette démonstration [ROBERTSON ET VIGNAUX, 1995]. De plus, la répétition d'un même énoncé par le même locuteur varie d'un énoncé à l'autre. L'existence d'une variabilité intralocuteur, aussi bien qu'une variabilité interlocuteur, implique que le processus inductif de détermination et de comparaison de l'identité de deux voix renferme lui-même une incertitude, associée à ce phénomène de variation [LEWIS, 1984].

1.3. Le rôle des probabilités dans l'identification

1.3.1. Définitions

La théorie des probabilités donne des règles qui permettent de déduire certaines probabilités inconnues, d'autres supposées connues et qui sont liées aux premières [KOLMOGOROV, 1933 *IN* : MATALON, 1967]. Selon l'école « subjectiviste », il est possible, en principe, pour une personne parfaitement cohérente de reconstituer objectivement les probabilités subjectives qu'elle attache à chaque événement. Évidemment dans cette perspective, la probabilité est une propriété d'un individu, et non d'un événement, comme c'est le cas pour l'école « objectiviste » ou « fréquentiste ».

Selon cette seconde école, il est possible de donner un statut objectif et non équivoque à la notion de probabilité, lorsqu'il s'agit d'événements susceptibles de se produire plusieurs fois. Elle refuse tout sentiment d'incertitude autre que celui qui porte sur l'occurrence de tels événements,

c'est-à-dire susceptibles de se répéter dans des conditions identiques, ce qui permet aux énoncés probabilistes d'être vérifiés empiriquement [MATALON, 1967].

Le chercheur, comme homme d'action, ne dispose qu'exceptionnellement de la totalité des informations qui lui seraient utiles pour aboutir à une conclusion ferme ou prendre la meilleure décision. L'objet de l'inférence statistique est d'utiliser au maximum l'information incomplète dont on dispose. Mais le saut qu'implique ce passage, de prémisses insuffisantes à une conclusion ou à une décision, ne peut se faire sans recours à un principe par nature extrinsèque ; l'information est ce qu'elle est, et aucune transformation tautologique ne pourra lui faire dire plus. Le principe d'inférence utilisé ne pourra donc jamais échapper totalement à l'arbitraire et est nécessairement dépendant de la conception qu'on a de la nature de la connaissance [MATALON, 1967].

Dans le domaine de l'inférence de l'identité, les deux types de probabilités, subjectives et objectives, sont utilisés, fait qui n'est pas toujours reconnu [LEWIS, 1984].

1.3.2. Limites de l'approche subjective

L'approche subjective rend possible la prise en compte d'une multitude de détails imparfaitement reconnus et difficiles à définir ou à cataloguer, impossibles à appréhender statistiquement [FINKELSTEIN ET FAIRLEY, 1970]. Pour la voix humaine, DODDINGTON définit cette information comme l'« information de haut niveau » véhiculée dans la parole [DODDINGTON, 1985].

La critique principale de l'approche subjective porte sur la difficulté à mesurer effectivement les probabilités personnelles. En effet, la théorie démontre l'existence, chez un individu rationnel, de probabilités qui traduisent ses degrés de croyance ; mais elle n'assure pas que les degrés de croyance exprimés directement par les individus réels sont bien des probabilités. Autrement dit, rien ne prouve que, lorsqu'un individu donne sa probabilité subjective de l'occurrence d'un événement ou de la vérité d'une proposition, cette grandeur ne satisfasse au calcul des probabilités [MATALON, 1967].

Un exemple frappant des dangers liés à la probabilité subjective a été donné par EVETT :

« Lors d'un jeu télévisé américain, un joueur se trouve face à trois rideaux (A, B et C), derrière l'un desquels se trouve un prix qu'il peut gagner, s'il fait le bon choix. Le présentateur connaît le rideau gagnant et invite le joueur à faire un premier choix, sans toutefois lui permettre de voir s'il a immédiatement gagné. À ce stade du jeu, le joueur a une chance sur trois de gagner.

Admettons que le joueur choisisse le rideau A, le présentateur montre alors que l'un des deux rideaux restants, par exemple C, ne cache pas le prix. À ce moment, le joueur est amené à faire un deuxième choix ; rester sur son premier choix, le rideau A, ou changer de rideau en faveur du rideau B.

La question est de savoir si le joueur a intérêt ou non à changer de rideau. Le fait de changer de rideau influence-t-il ses chances de gain ?

Intuitivement, on peut raisonner de la manière suivante : Une fois le rideau C soulevé par le présentateur, il ne reste que deux solutions : le prix est soit derrière A, soit derrière B. Que le joueur change ou non de rideau, ses chances de gain restent identiques : il n'y a donc aucun intérêt à changer lors du deuxième choix » [EVETT, 1992].

En réalité, le calcul des probabilités statistiques et la pratique répétée du jeu ont démontré que le joueur double ses chances de gain s'il change de rideau lors du deuxième choix. Ce jeu télévisé a suscité un large débat parmi les mathématiciens, la majorité d'entre eux tombant dans le piège de l'intuition. La solution correcte fut présentée par un profane [TIERNY, 1991 ; ENGEL ET VENETOULIAS, 1991].

1.3.3. Limites de l'approche statistique

L'approche statistique est limitée à l'opérationnalisation de variables objectivement mesurables et ne peut espérer sauvegarder la richesse de l'information analysée dans une procédure de reconnaissance ordinaire ou savante. DODDINGTON définit cette information véhiculée dans la parole humaine comme « information de bas niveau » [DODDINGTON, 1985].

Pour ces raisons, l'inférence statistique de l'identité est généralement plus faible que le jugement usuellement exprimé par un expert. Elle n'aboutit en fait que rarement à la démonstration de l'unicité des caractéristiques d'un indice matériel [FINKELSTEIN ET FAIRLEY, 1970]. Par contre, cette approche établit empiriquement la fréquence relative des caractéristiques étudiées, dans la limite où l'échantillon dans lequel elles sont observées représente la population générale [LEWIS, 1984].

1.4. Hypothèse de la recherche

L'hypothèse principale qui sous-tend cette recherche pose qu'il est possible d'extraire et d'analyser l'information dépendante du locuteur contenue dans la voix dans un but d'identification forensique.

1.5. Objectifs de la recherche

La connaissance de l'information dépendante du locuteur est empêchée par la difficulté à décrire symboliquement cette information, la procédure d'identification de locuteurs s'appuie sur une reconnaissance de cette information, sans qu'il soit possible de définir l'information elle-même [WARFEL, 1979 IN : THEVENAZ, 1993]. Selon HECKER, la reconnaissance de locuteurs se divise en trois divisions majeures : la reconnaissance de locuteurs par audition, par comparaison visuelle de spectrogrammes et par machine ou automatique [HECKER, 1971]. Les trois approches sont pratiquées en sciences forensiques, mais leurs performances sont actuellement mal définies dans le cadre de cette application. De plus, aucune n'obtient l'approbation de la communauté scientifique pour une utilisation en sciences forensiques, notamment parce que les processus d'inférence de l'identité du locuteur mis en œuvre ne sont pas satisfaisants.

Le premier objectif de cette recherche consiste à choisir un processus d'inférence de l'identité en conformité avec les exigences logiques et légales, de manière à mieux définir le rôle du raisonnement inductif et déductif dans les approches auditive, spectrographique et automatique pratiquées en sciences forensiques.

Les machines qui reconnaissent des mots dans la parole, ou des objets dans une image, sont investies d'une capacité d'identification. Il est cependant improbable que des algorithmes de reconnaissance de formes puissent disputer la suprématie à l'être humain dans la plupart des tâches. Par contre, dans certains domaines particuliers comme l'identification de personnes par l'écriture manuscrite ou par la voix, il est permis de douter des capacités de la perception humaine [LEWIS, 1984].

Le développement technologique rend possible la détection de traces et l'analyse de caractéristiques que personne n'avait pu effectuer auparavant et conduit à l'éclosion de nouvelles applications forensiques [ROBERTSON ET VIGNAUX, 1995]. Le progrès réalisé dans le domaine de l'intelligence artificielle perceptive, allié à l'accroissement des performances de calcul et de stockage des systèmes micro-informatiques, ouvre des perspectives d'application de la reconnaissance automatique de locuteurs aux sciences forensiques.

« Should a machine be produced that for a sample of speakers could be shown to perform the speaker identification task better than human listeners ? ... Clearly to make such a comparison it is necessary to define quantitatively what is meant by better performance for both man and machine » [LEWIS, 1984].¹

Le second objectif de cette recherche réside dans une tentative de réponse à la question de LEWIS, en évaluant les limites des probabilités subjectives et statistiques dans le processus de reconnaissance de locuteurs.

Une réponse à la question de LEWIS est même devenue urgente, puisque la recherche dans le domaine de la reconnaissance de locuteurs en sciences forensiques tend actuellement à se focaliser sur des procédures automatiques ou semi-automatiques, comme le souligne BRAUN en 1998, dans son rapport « *Voice Analysis* » présenté au congrès de l'Interpol [BRAUN, 1998].

1.6. Contributions majeures

Les contributions majeures de cette recherche peuvent être résumées ainsi :

- La démonstration de la conformité logique et légale d'un processus d'inférence de l'identité du locuteur dérivé du théorème de Bayes et basé sur une évaluation de rapports de vraisemblance.
- La détermination des limites des approches auditive, spectrographique et automatique dans leur application en sciences forensiques, sur la base d'une recherche bibliographique.
- La réalisation d'un système de reconnaissance automatique de locuteurs, reposant sur une technologie représentant l'état de l'art dans le domaine de la reconnaissance automatique de locuteurs indépendante du texte, et dont la structure permet une inférence de l'identité basée sur une évaluation de rapports de vraisemblance.

¹ Trad. : « Une machine devrait-elle être construite pour montrer qu'à partir d'un ensemble de locuteurs, elle réalise l'identification de locuteurs mieux que les auditeurs humains ? ... Clairement pour faire une telle comparaison, il est nécessaire de définir quantitativement ce qui est entendu par meilleure performance, tant pour l'homme que pour la machine ».

- L'enregistrement d'une base de données en langue française, répondant aux critères forensiques. Malgré sa taille modeste, sa structure a été définie de manière à bénéficier de la synergie d'une base de données de grande taille existante, pour l'évaluation de la variabilité interlocuteur. Son contenu est spécialement adapté à une utilisation forensique, avec la présence de locuteurs ayant des voix auditivement proches et de simulations d'indices matériels qui peuvent être rencontrées en cas d'abus de téléphone ou de mesure de surveillance.
- La mise au point d'un programme d'évaluation du système de reconnaissance développé, premièrement pour circonscrire au mieux la procédure nécessaire à l'obtention d'une évaluation réaliste de la variabilité intralocuteur et interlocuteur dans le contexte forensique et, deuxièmement, pour cerner les limites d'application du système dans un contexte forensique réel.

1.7. Organisation de la recherche

La première partie est consacrée à l'approche théorique de la reconnaissance de locuteurs en sciences forensiques. Après cette introduction, le deuxième chapitre concerne l'étude de la voix comme indice matériel en sciences forensiques. Le troisième chapitre s'attache à définir le rôle de l'expert en sciences forensiques, à déterminer les exigences légales en matière de preuve scientifique, à découvrir les différentes méthodes de reconnaissance de locuteurs, à analyser les processus d'inférence décrits pour l'identification du locuteur en sciences forensiques et à choisir le processus le plus approprié.

La deuxième partie a pour objet la recherche bibliographique. Le chapitre quatre cherche à déterminer les performances de l'être humain, profane ou expert, dans la tâche de reconnaissance de locuteurs ; le chapitre cinq étudie la méthode de reconnaissance de locuteurs par comparaison visuelle de spectrogrammes, alors que le chapitre six traite de l'approche automatique de la reconnaissance de locuteurs.

La troisième partie rend compte de la partie expérimentale de cette recherche, sur la base de l'approche théorique développée dans la première partie. Le chapitre sept expose les trois étapes du système de reconnaissance automatique développé : l'analyse du signal de parole, la classification et, plus en détail, le processus d'inférence de l'identité du locuteur. Le huitième chapitre décrit la réalisation de la base de données « Polyphone IPSC » et livre les résultats de l'évaluation du système de reconnaissance automatique développé.

La dernière partie est une discussion générale. Elle comporte une partie rétrospective relative à la présente recherche, et une partie prospective qui traite de l'utilisation du système développé dans un contexte réel et des problèmes non résolus. Elle se termine par une conclusion en forme de synthèse.

II. LA VOIX COMME INDICE MATERIEL

2.1. Introduction

Dans toute enquête comprenant des enregistrements de parole, l'écoute de l'indice enregistré constitue la tâche initiale et la seule, lorsque les personnes chargées de l'enquête ne peuvent établir aucune relation entre la voix inconnue et une personne connue [BOLT *ET AL.*, 1979]. Cet examen préliminaire implique que le recours à une expertise en reconnaissance de locuteurs a lieu le plus souvent lorsqu'une ressemblance auditive frappante est constatée entre l'enregistrement de parole inconnue et la voix d'une personne mise en cause, mais que celle-ci nie, ou lorsque la présomption de déguisement existe sur la base de l'écoute préliminaire [NOLAN, 1991 ; BRAUN, 1994 ; BROEDERS, 1995] (Figure II.1).

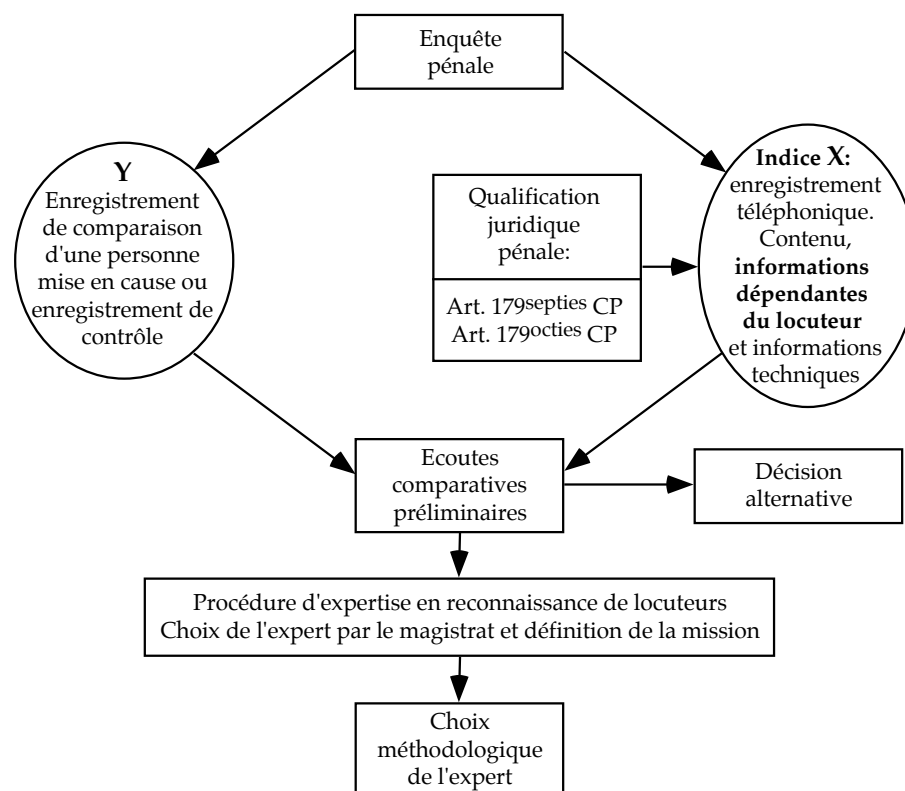


Figure II.1. Place de la procédure d'expertise en reconnaissance de locuteurs dans l'enquête pénale

La plupart des indices sont des enregistrements qui résultent d'écoutes téléphoniques ou de messages anonymes ; en Allemagne, c'est par exemple le cas pour 95% des échantillons de voix inconnue analysés [KÜNZEL, 1994A]. Cette particularité, concernant le mode de collecte de l'indice matériel, mérite une présentation du cadre légal et des aspects techniques qui entourent actuellement la procédure d'enregistrement en Suisse.

2.2. Cadre légal

La protection de la sphère privée, et notamment la protection des relations établies par les télécommunications, est considérée comme un droit fondamental en Suisse. La législation concernant l'autorisation, pour l'État, de procéder à des écoutes téléphoniques est donc politiquement très délicate et souvent remise en question dans les États de droit. Preuve en est l'intensité des débats du Conseil National suisse à l'automne 1996, au sujet de l'initiative populaire et la loi fédérale « S.o.S. Pour une Suisse sans police fouineuse. Maintien de la sûreté intérieure », ou la mobilisation de l'opposition japonaise pour empêcher le Sénat nippon d'adopter une loi autorisant les écoutes téléphoniques, votée au début juin 1999 par les députés de la Chambre des représentants [HAYANO, 1999].

2.2.1. Conditions de recevabilité d'un enregistrement téléphonique en Suisse

Le secret de la correspondance téléphonique est un droit fondamental garanti par l'art. 36 al. 4 de la Constitution fédérale de la Confédération suisse (CF)² du 29 mai 1874), et repris dans l'art. 13 al. 1 de la mise à jour de la Constitution fédérale, proposée par l'Assemblée fédérale du 18 décembre 1998, approuvée par le peuple le 19 avril 1999, et qui est entrée en vigueur le 1^{er} janvier de l'an 2000. En bref, un inculpé ou un suspect peut être placé sous surveillance si les trois conditions suivantes sont réalisées :

- La poursuite pénale concerne un crime ou un délit dont la gravité ou la particularité justifie l'intervention, ou une infraction quelconque commise au moyen du téléphone, même une simple contravention, comme l'abus de téléphone au sens de l'art. 179^{septies} du Code pénal suisse (CP)³.
- La personne mise sous surveillance est soupçonnée, en raison de faits déterminés, d'avoir commis l'infraction ou d'avoir participé à sa perpétration.
- À défaut de surveillance, les investigations nécessaires étaient notablement plus difficiles à mener ou d'autres actes d'instruction n'ont pas permis d'obtenir de résultat [GAUTHIER, 1984].

En droit pénal, l'enregistrement non autorisé d'une conversation non publique est un délit poursuivi sur plainte et passible de l'emprisonnement ou de l'amende, selon l'art. 179^{ter} al. 1 CP. La surveillance officielle est justifiée par l'art. 179^{octies} al. 1 CP. L'écoute et l'enregistrement par des particuliers peuvent l'être pour la légitime défense ou l'état de nécessité, respectivement art. 33 et art. 34 CP [STRATENWERTH, 1983].

La plupart des lois ne définissent pas les modes de preuve et laissent le juge du fait libre de former son intime conviction sur tous les éléments apportés par l'instruction. L'enregistrement sonore clandestin des paroles d'autrui par un particulier n'est donc pas, en soi, inapte à servir de preuve, s'il est démontré qu'il est fidèle et qu'il n'a pas été modifié. Les circonstances dans

² *infra* : Annexe I. Extraits de la Constitution fédérale de la Confédération suisse

³ *infra* : Annexe II. Extraits du Code pénal suisse

lesquelles les propos ont été enregistrés doivent être élucidées et ce moyen de preuve doit être écarté s'il porte atteinte aux droits de la personnalité [GAUTHIER, 1984].

Quant à l'enregistrement effectué au su d'un interlocuteur, mais contre son gré, les avis sont partagés. D'aucuns pensent que celui qui s'est opposé à l'enregistrement de ses déclarations, mais parle néanmoins, sachant que ses propos sont enregistrés, donne son consentement par actes concluants [STRATENWERTH, 1983]. Pour d'autres au contraire, l'enregistrement n'est pas autorisé et demeure punissable [SCHULTZ, 1971].

Si un accord écrit de la part des personnes enregistrées, lors de procédures d'enregistrement de comparaison, peut permettre de lever cette incertitude juridique, la conscience d'être enregistré est susceptible d'influencer très négativement la constitution d'enregistrements représentatifs d'une élocution spontanée⁴. Le locuteur peut délibérément altérer son élocution par une stratégie de déguisement systématique, ou, s'il est coopératif, le stress ou la peur engendrée par cette procédure peut induire une modification involontaire de son élocution [BROEDERS, 1995].

2.2.2. La procédure d'écoute téléphonique en Suisse

La procédure d'écoute téléphonique est soumise à l'ordonnance sur le service de surveillance de la correspondance postale et des télécommunications du 1^{er} décembre 1997⁵. La surveillance est ordonnée par un magistrat de l'ordre exécutif ou judiciaire, désigné par la loi, et est effectuée par le prestataire de service de télécommunication à la demande du service fédéral de la surveillance de la Poste et des télécommunications.

Les enregistrements qui ne sont pas nécessaires pour l'enquête sont conservés sous clé et détruits à l'issue de la procédure. Ils ne sauraient être conservés plus longtemps au titre de pièces à conviction auxiliaires, ni utilisés dans une autre procédure sans être soumis à la consultation des parties [GAUTHIER, 1984 ; PIQUEREZ, 1994].

L'avènement de la technologie des réseaux cellulaires numériques rend considérablement plus difficile la procédure d'écoute. En effet, s'il est possible de déterminer facilement le numéro d'appel d'un téléphone portable acheté avec un abonnement, la situation d'un appareil acheté avec une carte à prépaiement ou d'un appareil volé est très différente. Leurs numéros sont uniquement accessibles aux systèmes de scanners électroniques, qui interceptent la transmission des codes d'identification des appareils, à des fins de facturation, et l'éventuel numéro d'identification personnel de l'utilisateur [MCCULLEY ET RAPPAPORT, 1993 IN : NATARAJAN ET AL., 1995].

Malheureusement, ce type de scanner permet aussi à des utilisateurs illégitimes d'intercepter ces codes, pour programmer des appareils clones d'appareils légitimes, leur offrant ainsi la possibilité d'une utilisation illicite virtuellement illimitée et sans grands risques [NATARAJAN ET AL., 1995]. Cette procédure est d'autant plus facile que les logiciels informatiques et les instructions de reprogrammation des téléphones portables sont disponibles sur le réseau informatique Internet

⁴ *infra* : 2.3.8. Influence du locuteur

⁵ *infra* : Annexe III. Ordonnance sur le service de surveillance de la correspondance postale et des télécommunications

[GAURA, 1994 IN : NATARAJAN ET AL., 1995]. Comme contre-mesure, NATARAJAN mentionne la mise en place de systèmes sommaires d'identification de locuteurs capables de détecter la présence d'une voix inhabituelle, mais l'efficacité de cette méthode n'est pas connue [NATARAJAN ET AL., 1995].

La sécurité devrait être améliorée par la technologie de cryptage numérique et par celle qui consiste à modifier les codes d'identification à chaque appel (*tumbling*). Par contre, ces mesures devraient rendre très difficile, voire impossible, l'écoute téléphonique de communications et l'identification d'appareils équipés de ce système [DE MARIA, 1994].

Finalement, l'arrivée sur le marché d'appareils téléphoniques portables dotés de méthodes cryptographiques robustes (*strong encryption*) du signal de parole rend obsolète l'art. 7 al. 2 de l'ordonnance sur le service de surveillance de la correspondance postale et des télécommunications du 1^{er} décembre 1997⁶. Celle-ci stipule que :

« les fournisseurs de service de télécommunication fournissent dans les meilleurs délais les relevés de service demandés et transmettent si possible en temps réel les communications de la personne surveillée. Ils suppriment les cryptages ».

Cette dernière phrase n'est pas satisfaisante, puisqu'elle n'est pas applicable dans les faits. En effet, c'est le constructeur du matériel qui est à l'origine de la mise à disposition de la technologie de cryptage et non le prestataire de services de télécommunication, et comme aucun moyen technique ne permet de supprimer ce cryptage, la surveillance téléphonique devient parfaitement impossible. Cette disposition risque même d'être contre-productive, en laissant croire aux autorités suisses chargées de la répression des infractions que les problèmes liés au décryptage et au déchiffrement des méthodes cryptographiques robustes sont maîtrisés et résolus, alors qu'il n'en est rien [MEUWLY, 1999].

2.3. Collecte de l'indice matériel

L'indice matériel ne consiste pas en la voix elle-même mais en un enregistrement téléphonique, c'est-à-dire en une transposition obtenue par un transducteur, qui convertit l'énergie acoustique en une autre forme d'énergie : mécanique, électrique ou magnétique. Cette information convertie est encodée par une méthode de codage analogique ou numérique, transmise par un réseau téléphonique et enregistrée dans une mémoire de masse. Dans un enregistrement analogique la puissance et la forme de l'onde sont en relation directe avec l'onde acoustique originale. Dans un enregistrement numérique l'onde est transposée et échantillonnée ; ensuite chaque échantillon est converti en un nombre binaire.

En Suisse, la parole est essentiellement transmise de manière numérique dans les réseaux téléphoniques, commutés et cellulaires ; par contre la technologie d'enregistrement mise en œuvre par l'État est encore essentiellement analogique. Une analyse de tous les éléments de la chaîne conduisant à l'enregistrement de l'indice matériel, du locuteur au système d'enregistrement,

⁶ *infra* : Annexe III. Ordonnance sur le service de surveillance de la correspondance postale et des télécommunications

permet de mieux évaluer l'influence de chaque maillon dans l'obtention de la qualité finale de l'indice.

2.3.1. Description et représentation du signal de parole

2.3.1.1. Structure

Le signal de parole est un signal réel non stationnaire, continu et d'énergie finie. Sa structure est complexe et variable dans le temps, pseudo-périodique pour les sons voisés, aléatoire pour les sons fricatifs et impulsionnelle dans les phases explosives des sons occlusifs (Figure II.2.).

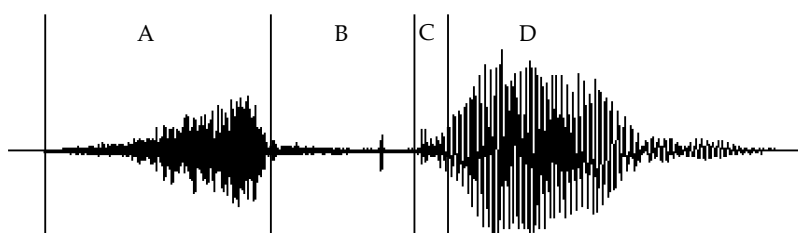


Figure II.2. La forme d'onde du signal vocal /ski/ : A. aléatoire – B. bruit – C. impulsionnelle – D. pseudo-périodique

2.3.1.2. Représentation numérique

Le signal acoustique temporel continu n'est accessible aux techniques numériques de traitement du signal que sous une représentation échantillonnée, quantifiée et limitée en temps [DRYGAJLO, 1999].

2.3.1.2.1. Échantillonnage

La représentation numérique implique un échantillonnage du signal effectué à une fréquence f_e compatible avec les exigences du théorème de Shannon. Selon ce théorème, la perte d'information entre le signal temporel continu et le signal discret correspondant est nulle si et seulement si la fréquence d'échantillonnage est au moins supérieure ou égale au double de la fréquence la plus haute contenue dans ce signal, appelée fréquence de Nyquist.

2.3.1.2.2. Quantification

Chaque échantillon est quantifié avec un pas de quantification q en rapport avec la précision souhaitée et codé par un algorithme qui dépend de la nature et des exigences de l'application. Pour un convertisseur analogique-numérique où n représente le nombre de bits des valeurs de sortie, le rapport signal sur distorsion de quantification, mesuré en dB, varie linéairement avec n et augmente de 6 dB avec chaque bit supplémentaire ; le niveau de la distorsion de quantification dépend de la composition fréquentielle du signal [DE COULON, 1990].

2.3.1.2.3. Codage

Le codage est réalisé par des méthodes temporelles, paramétriques ou hybrides. Les méthodes de codage temporelles cherchent à approximer le signal de parole en tant que forme

d'onde et sont utilisées avec une quantification uniforme, lorsqu'une haute fidélité de restitution du signal est nécessaire.

Les méthodes paramétriques cherchent plutôt à modéliser le processus de production de la parole et à extraire les paramètres pertinents qui sont transmis au décodeur. Ceux-ci permettent de reconstruire une forme d'onde souvent éloignée de la forme du signal initial, mais qui produit un son subjectivement proche de l'original. L'utilisation de méthodes purement paramétriques est actuellement très limitée car elles entraînent une dégradation trop importante du naturel de la voix et une sensibilité excessive à l'influence de l'environnement de la prise de son. Seules les méthodes hybrides, qui font intervenir à la fois les méthodes temporelles et les méthodes paramétriques, sont actuellement capables de fournir des résultats satisfaisants dans des applications nécessitant de fortes réductions de débit.

2.3.2. Mesure de la qualité de la parole

La mesure de la qualité de la parole transmise par un système technique est délicate car elle est subjective et dépend de nombreux paramètres. Plusieurs échelles permettent d'évaluer globalement cette qualité, mais les paramètres sur lesquels ces échelles sont définies sont tous centrés sur le problème de la qualité subjective de la perception, bien qu'un système de communication vocal doive non seulement garantir un confort de perception, mais aussi préserver les caractéristiques permettant l'identification de chaque voix [SCHMIDT-NIELSEN ET STERN ; 1985].

Le *Mean Opinion Score* (MOS) est une échelle graduée sur cinq niveaux : 1 = mauvais (*bad*) ; 2 = médiocre (*poor*) ; 3 = suffisant (*fair*) ; 4 = bon (*good*) et 5 = excellent (*excellent*) [DAUMER, 1982]. Des scores plus élevés que 4.0 correspondent à une qualité élevée ou à un système de codage du signal presque transparent. La qualité théorique du réseau téléphonique public commuté se situe entre 4.0 et 4.5, mais, pour que cette qualité soit atteinte, tous les éléments du réseau doivent se trouver à ce niveau. De fait, la qualité réelle des réseaux est toujours en deçà de cette qualité théorique, mais cette information n'est pas disponible, car elle fait partie des données sensibles des entreprises de télécommunication. Des scores entre 3.5 et 4.0 correspondent à une qualité de communication de téléphonie cellulaire ou de synthèse vocale. Ce niveau de qualité est caractérisé par une dégradation aisément détectable, sans pour autant gêner la transmission téléphonique naturelle. La parole de qualité synthétique, utilisée dans les systèmes de transmission militaires, atteint des scores qui n'excèdent pas 3.0. Elle peut impliquer un signal d'intelligibilité élevée, mais trop peu naturel pour permettre la reconnaissance auditive de locuteurs.

Le *Diagnostic Rhyme Test* (DRT) est une mesure de l'intelligibilité des mots, alors que la *Diagnostic Acceptability Measure* (DAM) reflète l'acceptabilité générale de la communication parlée. Le résultat de ces deux tests est exprimé en pour cent [VOIERS, 1977A ; VOIERS, 1977B].

2.3.3. Influence du système de codage numérique de l'information

L'aire d'audition humaine est comprise entre le seuil d'audition, qui varie entre 0 et 40 dB selon la fréquence, et le seuil de la douleur, situé aux alentours de 120 dB. Dans le domaine fréquentiel, la sensibilité de l'oreille s'étend entre 16 Hz et 20 kHz.

2.3.3.1. Haute fidélité

Pour couvrir entièrement cette aire d'audition dans le domaine numérique et obtenir une haute fidélité de restitution, les convertisseurs analogiques numériques exploitent des systèmes de codage de type *Pulse Code Modulation* (PCM). Le signal est échantillonné à une fréquence de 48 kHz pour les systèmes d'enregistrement professionnels comme le *Digital Audio Stationery Head* (DASH) ou le *Digital Audio Tape* (DAT), de 44,1 kHz pour le *Compact Disc* (CD) et 32 kHz pour le *Digital Audio Broadcasting* (DAB), ce qui assure une bande passante du signal d'au moins 16 kHz.

Les convertisseurs analogiques numériques professionnels quantifient les échantillons sur 24 bits, ce qui correspond à une résolution de $16'777'216$ niveaux (2^{24}) et représente un rapport signal sur distorsion de quantification de l'ordre de 140 dB. Pour les convertisseurs haute fidélité grand public le codage est généralement réalisé sur 16 bits, ce qui correspond à une résolution de $65'536$ niveaux (2^{16}) et à un rapport signal sur bruit de l'ordre de 93 dB. Le DAB se contente d'une résolution non uniforme sur 12 bits. L'obtention de telles qualités sonores nécessite un débit binaire compris entre 384 kbits s^{-1} ($12 \text{ bits} * 32 \text{ kHz}$) et $1152 \text{ kbits s}^{-1}$ ($24 \text{ bits} * 48 \text{ kHz}$) par canal.

2.3.3.2. Réseau téléphonique public commuté (RTPC)

Les systèmes de codage développés pour la téléphonie permettent une forte réduction du débit binaire de l'information, mais impliquent une perte de qualité perceptible (Figure II.3.).

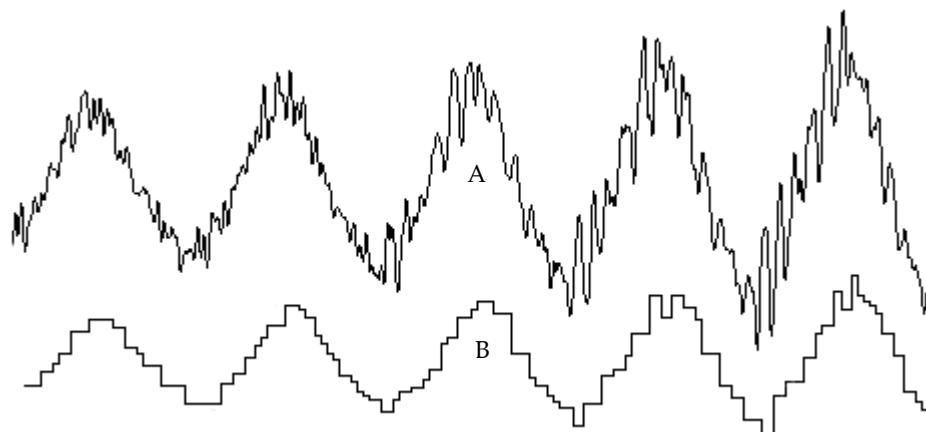


Figure II.3. Extrait d'un centième de seconde du phonème /i/ signal vocal /ski/ : A. 16 bits-44.1 kHz B. 8 bits-8 kHz

Pour le réseau téléphonique public commuté (RTCP), *Public Switched Telephone Network* (PSTN), la première norme de codage numérique a été recommandée en 1972 par l'ITU-T (*International Telephone Union - Telephony division*). Baptisée G.711, elle est basée sur l'algorithme PCM ; le signal est numérisé avec une fréquence d'échantillonnage de 8 kHz et quantifié de manière non uniforme sur 8 bits, selon la loi A en Europe et la loi μ aux États-Unis, dont les qualités sont équivalentes. Ce débit binaire de 64 kbits s^{-1} assure une bande passante de 300 à 3400 Hz. Les normes G.721 et G.726, définies en 1984 et 1990, sont basées sur l'algorithme *Adaptive Differential Pulse Code Modulation* (ADPCM), qui ne code plus directement l'amplitude de

l'échantillon, mais seulement la différence entre l'amplitude et une valeur prédite par un filtrage de type adaptatif, ce qui permet une réduction du débit binaire à 32 kbits s⁻¹ (G.721) ou un débit variable de 40, 32, 24 et 16 kbits s⁻¹ (G.726). En 1991, l'ITU-T a sélectionné un système de codage à 16 kbits s⁻¹ (G.728), basé sur une technique hybride de modélisation et de quantification vectorielle⁷, l'algorithme *Low-Delay Code Excited Linear Prediction Coder* (LD-CELP), qui présente un faible délai de reconstruction, propriété particulièrement importante pour un échange téléphonique (Tableau II.1.).

Le standard G.729, défini par l'ITU-T pour le système de codage à 8 kbits s⁻¹ est un compromis entre l'algorithme *Algebraic Code Excited Linear Prediction Coder* (ACELP) présenté par l'Université de Sherbrooke au Canada en association avec France Telecom et l'algorithme *Conjugate Structure Code Excited Linear Prediction Coder*, proposé par *Nippon Telephone & Telegraph* (NTT) au Japon [KONDOZ, 1994 ; DRYGAJLO, 1999].

2.3.3.3. Réseau téléphonique cellulaire

En Europe, la norme *Global System for Mobile communication* (GSM) a été définie en 1989 et la norme *Cellular Telecommunication Industry Association* (CTIA) IS-54 a été définie pour l'Amérique du Nord en 1990. La première génération de systèmes de codage (*Full Rate*) fait appel à des techniques d'accès multiples par division de temps (TDMA) et à un codeur de source à 13 kbits s⁻¹ *Regular Pulse Excitation - Long Term Prediction* (RPE-LTP) pour l'Europe et *Vector Sum Excited Linear Prediction* (VSELP) à 8 kbits s⁻¹ pour l'Amérique du Nord. Tous les débits mentionnés font référence au codage de source ; le codage du canal utilise approximativement le même débit, ce qui porte le débit total à 22,8 kbits s⁻¹ pour le GSM. Cette première génération ne permet qu'une multiplication par trois environ des capacités de ce réseau par rapport au réseau analogique. Pour permettre une multiplication par dix ou plus, l'*European Telecommunication Standard Institute* (ETSI) en Europe et la CTIA en Amérique du Nord choisissent actuellement les standards de la deuxième génération. Le nouveau système de codage *GSM Half Rate* (HR) est basé sur un algorithme de type *Code Excited Linear Prediction Coder* (CELP), de débit binaire de 5,6 kbits s⁻¹, alors que le système *GSM Enhanced Full Rate* (EFR) est basé sur un algorithme de type CELP, de débit binaire de 12,2 kbits s⁻¹.

Les systèmes de communication mobile de troisième génération, dont le débit devrait atteindre plusieurs Mbits par seconde, *Universal Mobile Transmission System* (UMTS) en Europe et *Future Public Land Mobile Telecommunication System* (FPLMTS) en Amérique du Nord, devraient être basés sur les techniques d'accès multiples par division de code (CDMA). Cette technique permettrait, entre autres avantages, une réalisation naturelle du débit variable. La CTIA a d'ailleurs standardisé l'algorithme *Qualcomm Code Excited Linear Prediction Coder* (QCELP) de la société Qualcomm® (IS-95), qui sélectionne dynamiquement toutes les 20 ms un débit de 8, 4, 2 ou 1 kbits s⁻¹ ; le débit moyen est de l'ordre de 4 kbits s⁻¹ [KONDOZ, 1994 ; DRYGAJLO, 1999] (Tableau II.1.).

⁷ *infra* : 6.3.2.2. Représentation par quantification vectorielle

2.3.3.4. Communications sécurisées et communications par satellite

Pour les réseaux de communication par satellite, les communications militaires comme celles de l'Organisation du Traité de l'Atlantique Nord (OTAN) et les communications intergouvernementales sécurisées, les débits binaires sont encore plus faibles. Le choix du maintien de l'intelligibilité se fait au détriment des caractéristiques dépendantes du locuteur, ce qui n'est pas sans poser certaines questions sur l'identification des interlocuteurs. Le système de codage exploité par le réseau satellitaire INMARSAT (*International Maritime Satellite*) est basé sur l'algorithme *Multi-Band Excitation* (MBE), dont le débit binaire est de 16 kbits s⁻¹. Les communications militaires de l'armée des États-Unis, de l'OTAN et de l'Armée Suisse utilisent un système de codage FS-1016, standardisé par le département de la Défense des États-Unis (USDoD) et basé sur un algorithme CELP à 4,8 kbits s⁻¹ [KONDOZ, 1994 ; DRYGAJLO, 1999] (Tableau II.1.).

Type de réseau	Standard de codage numérique	Taux de transfert (kbits s ⁻¹)	Système de codage	MOS	DRT	DAM
Téléphonique commuté	G.711 (1 ^{ère} génération)	64	<i>Pulse Code Modulation</i> (PCM)	4,3	95	73
Téléphonique commuté	G.721 (2 ^{ème} génération)	32	<i>Adaptive Differential Pulse Code Modulation</i> (ADPCM)	4,1	94	68
Téléphonique commuté	G.726 (2 ^{ème} génération)	16 - 40	<i>Code Excited Linear Predictive Coder</i> (CELP)	-	-	-
Téléphonique commuté	G.728 (3 ^{ème} génération)	16	<i>Code Excited Linear Predictive Coder</i> (CELP)	4,0	94	70
Téléphonique satellitaire	INMARSAT Standard B	16	<i>Muti-Band Excitation</i> (MBE)	-	-	-
Téléphonique cellulaire (Europe)	GSM (1 ^{ère} génération)	13	<i>Regular Pulse Excitation - Long Term Prediction</i> (RPE-LTP)	-	-	-
Téléphonique cellulaire (USA)	CTIA IS-54 (1 ^{ère} génération)	8	<i>Vector Sum Excited Linear Prediction</i> (VSELP)	3,7	93	68
Téléphonique cellulaire (Europe)	GSM <i>Enhanced Full Rate</i> (EFR) (2 ^{ème} génération)	12,2	<i>Code Excited Linear Predictive Coder</i> (CELP)	-	-	-
Téléphonique cellulaire (Europe)	GSM <i>Half Rate</i> (HR) (2 ^{ème} génération)	5,6	<i>Code Excited Linear Predictive Coder</i> (CELP)	-	-	-
Militaire	DOD-CELP (FS 1016)	4,8	<i>Code Excited Linear Predictive Coder</i> (CELP)	3,0	93	67
Militaire	LPC-10 (FS 1015)	2,4	<i>Linear Prediction Coder</i> (LPC) ⁸	2,5	90	54

Tableau II.1. Comparaison de la qualité du signal transmis par plusieurs systèmes de codage [JAYANT, 1992 ; KONDOZ, 1994]

⁸ *infra* : 6.2.3.1. Prédiction linéaire

2.3.4. Influence de la prise de son

La qualité de la prise de son dépend des caractéristiques du microphone (Figure II.4.) et des caractéristiques acoustiques de l'endroit où se trouve le locuteur, alors que son élocution dépend principalement de l'environnement sonore de cet endroit. La manifestation la plus connue de l'influence de l'environnement sur le locuteur est sans aucun doute l'adaptation de l'intensité de la voix au niveau sonore ambiant, « l'effet LOMBARD » [LOMBARD, 1911 IN : HATON, 1994]. La conjonction de toutes ces influences peut contaminer le message de multiples façons par des bruits de fond de convolution ou additifs, avant même sa transmission dans le réseau téléphonique [BOLT ET AL., 1979].

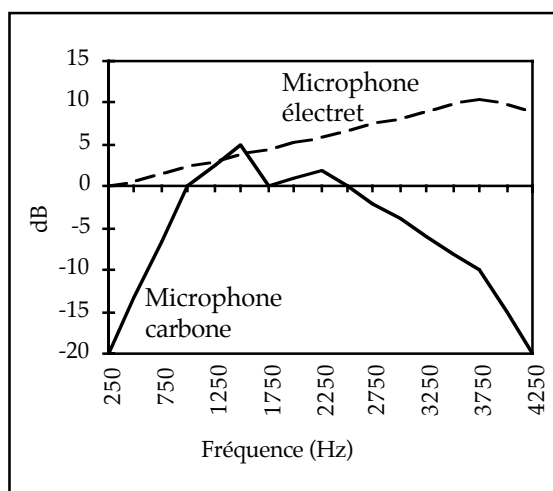


Figure II.4. Comparaison de la réponse en fréquence de deux types de microphones [HUNT, 1991]

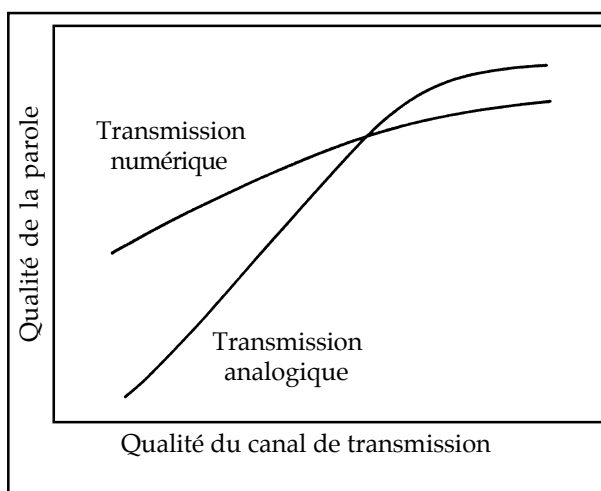


Figure II.5. Comparaison de la qualité de transmission de deux types de réseaux téléphoniques cellulaires [JAYANT, 1992]

2.3.5. Influence du canal de transmission

Dans la transmission téléphonique, les dégradations dépendent aussi du type de réseau téléphonique, fixe ou cellulaire, et du système de transmission des informations dans ce réseau, analogique, numérique ou combiné (Figure II.5.) [JAYANT, 1992].

Le réseau téléphonique suisse est desservi à plus de 90 % par des centraux numériques et cette évolution touche tous les pays économiquement développés [BROEDERS, 1995]. Par contre, lors de communications téléphoniques avec des pays dont les réseaux analogiques sont anciens ou de mauvaise qualité : la qualité de la voix de l'interlocuteur se trouvant dans l'un de ces pays peut être très inférieure à celle de l'interlocuteur se trouvant dans un pays développé. La qualité de la transmission dépend aussi du système de codage numérique utilisé, différent dans les réseaux commutés et cellulaires.

2.3.5.1. Communication sur le réseau téléphonique public commuté

En Suisse, la qualité du réseau téléphonique public commuté, entièrement numérique, est située entre 4,0 et 4,5 MOS ; bien qu'aucune évaluation précise ne soit disponible, il semble qu'il soit excellent, en comparaison internationale. Le réseau est organisé en étoile et la conversion analogique-numérique est effectuée dans le central de quartier, en cas de raccordement analogique, ou à l'intérieur même de l'appareil en cas de raccordement RNIS (Réseau Numérique à Intégration de Services). Ce dernier type de raccordement permet la transmission simultanée de deux canaux de voix et d'un canal de signalisation.

Dans le réseau téléphonique public commuté, le combiné des appareils est relié au téléphone soit par un fil, soit par une liaison hertzienne, pour les appareils dits « sans fil ». Pour les appareils « sans fil » analogiques, la transmission entre le téléphone et le combiné est assurée par modulation de fréquence à 27 MHz. Les appareils actuels exploitent une technologie numérique nommée DECT (*Digital European Cordless Telephone*), proche de celle du GSM : 120 canaux numériques sont distribués entre 1880 et 1900 MHz.

2.3.5.2. Communication sur le réseau téléphonique numérique cellulaire

Dans le domaine de la téléphonie numérique cellulaire, l'évolution des algorithmes de codage tend vers une diminution des taux de transfert, tout en garantissant une qualité de communication acceptable, supérieure ou égale à 3.0 MOS. La conversion analogique numérique est effectuée à l'intérieur de l'appareil émetteur-récepteur. L'appareil portable émet le signal numérique entre 890 et 915 MHz et la station de base entre 935 et 960 MHz [KREBSER, 1993]. Pour augmenter le nombre de raccordements simultanés au réseau, le réseau DCS 1800 (*Digital Cellular System*) est actuellement mis en place en Suisse. Dans ce second réseau, la station de base émet le signal entre 1805 et 1880 MHz et l'appareil portatif entre 1710 et 1785 MHz.

Les algorithmes de codage numériques sont optimisés pour la transmission du signal de la parole. La contamination de ce signal par du bruit de fond engendre des distorsions non linéaires, impossibles à modéliser analytiquement. Ce problème est surtout présent dans le cas de la téléphonie mobile, car les appareils cellulaires sont utilisés dans des environnements sonores très divers et bruyants, notamment en voiture.

2.3.5.3. Communication sur le réseau téléphonique par satellite

Une qualité de 3.0 MOS est acceptée pour une utilisation militaire et professionnelle extrême, comme celle des correspondants de guerre et des navigateurs, dans laquelle la sauvegarde de l'intelligibilité est l'essentiel. Par contre, la récente expérience du premier projet de couverture totale par satellite *Iridium*TM de *Motorola*[®] montre qu'elle ne satisfait pas le consommateur. En effet, outre un prix prohibitif et un appareil lourd et encombrant, la mauvaise qualité de transmission est le principal grief adressé par les utilisateurs à l'égard de ce système.

2.3.6. Influence du système d'enregistrement

Dans la procédure de collecte de l'indice matériel, la qualité du système d'enregistrement est le seul maillon de la chaîne qu'il est possible d'influencer, mais c'est malheureusement souvent le plus faible [BOLT ET AL., 1979].

2.3.6.1. Enregistrement dans le cadre d'une mesure de surveillance

Dans le cadre d'une mesure officielle de surveillance téléphonique, les enregistrements sont confiés aux prestataires de services de télécommunication, sur demande du service fédéral de la surveillance de la Poste et des télécommunications. La qualité de ces enregistrements pourrait être optimale dans les limites des caractéristiques de la prise de son et du canal de transmission, si des systèmes adéquats étaient mis en œuvre, ce qui n'est encore que très partiellement le cas en Suisse. Ceci nécessiterait d'une part un enregistrement direct de l'information numérique sans compression du signal et d'autre part la séparation des signaux provenant des différents interlocuteurs et l'enregistrement de chacune des voix sur une piste séparée en cas de dialogue ou de conversation entre plusieurs personnes. Cette dernière mesure éviterait toute procédure de ségrégation des locuteurs, manuelle ou automatique, et faciliterait grandement le travail de retranscription des conversations téléphoniques, en permettant une écoute indépendante des intervenants.

2.3.6.2. Enregistrement dans le cadre d'un abus de téléphone

Lorsque les abus de téléphone sont destinés à un service officiel dont les communications téléphoniques sont enregistrées, en particulier les services du feu, de police et de santé, les messages sont conservés sur des bandes magnétiques analogiques, dont la vitesse de défilement est très lente. La durée d'enregistrement obtenue est très grande, jusqu'à 24 heures, mais la qualité est très faible. Lorsque les abus de téléphone visent d'autres abonnés, les enregistrements proviennent en général de répondeurs téléphoniques, seules installations accessoires d'enregistrement autorisées par les prestataires de service de télécommunication. Si le message provient lui-même d'un enregistrement, la qualité en est encore amoindrie.

2.3.6.3. Standard de qualité en matière d'enregistrements téléphoniques

Actuellement, les systèmes d'acquisition et d'édition numérique assistés par ordinateur font presque partie des applications grand public. Le CD devient un support numérique quasi universel, tant pour les données audio qu'informatiques, et le prix de revient du support de type *Recordable Compact Disc* (CD-R) est largement inférieur à celui d'une cassette audio compact de qualité. Les forces de police de Suisse sont encore largement équipées de matériel d'enregistrement analogique obsolète, mais le moment est judicieux pour un passage à une stratégie d'enregistrement numérique, avec comme base des standards de qualité et des protocoles d'acquisition communs dans tout le pays.

2.3.7. Influence du type d'investigation

2.3.7.1. Influence de l'investigation préliminaire

Comme déjà dit en introduction, dans toute enquête comprenant des enregistrements de parole, l'écoute de l'enregistrement considéré comme indice constitue la tâche initiale et la seule, lorsqu'aucune relation avec une voix connue ne peut être établie par les personnes chargées de l'enquête, un témoin ou une victime [BOLT *ET AL.*, 1979]. Ce tri préliminaire implique que le recours à une analyse a lieu le plus souvent lorsqu'une ressemblance auditive frappante est constatée entre la voix présente sur l'indice et la voix d'une personne mise en cause ou lorsque la présomption de déguisement existe sur la base de cette écoute [NOLAN, 1991 ; BRAUN, 1994 ; BROEDERS, 1995].

Dans le canal téléphonique, l'intelligibilité et le taux d'identification des locuteurs décroissent en raison de la réduction de la bande passante et l'addition de bruit. Dans des conditions d'enregistrement de haute qualité, ces dégradations altèrent plus le taux d'identification et dans des conditions d'enregistrement de basse qualité, elles affectent plus l'intelligibilité [CLARKE *ET AL.* ; 1966]. Ces résultats laissent à penser que l'intelligibilité de la parole dégradée n'est pas un indicateur fiable pour l'identification, bien que ce critère soit probablement très utilisé dans la tâche initiale d'écoute et de tri des échantillons.

Par contre, la détermination subjective de la langue parlée peut être considérée comme fiable, si elle est réalisée par une personne qui maîtrise cette langue, puisque cette faculté fait alors partie du sens commun. Ceci peut néanmoins signifier la nécessité pour l'autorité policière ou judiciaire de recourir aux services d'un linguiste, d'un interprète ou d'un traducteur assermenté.

2.3.7.2. Enregistrement dans le cadre d'un abus de téléphone

Si le message est enregistré dans le cadre d'un abus de téléphone, au sens de l'art. 179^{septies} CP⁹, il est en général de courte durée, de quelques secondes à quelques minutes, et contient une faible quantité d'information. Son contenu peut constituer en lui-même une infraction, dont la qualification dépend de l'intention de son auteur et des thèmes abordés. La taille des champs lexicaux utilisés est en général restreinte, à cause de la brièveté de l'énoncé, et les thèmes sont limités : menaces, injures, propos obscènes, thèmes propres aux pathologies psychiatriques [FÄHRMAN, 1966A]. S'il n'y a pas d'échange entre les interlocuteurs, il est possible que le message provienne lui-même d'un enregistrement et qu'il ait été volontairement modifié au cours de cette procédure, soit par un filtrage, soit par un montage [BOLT *ET AL.*, 1979].

2.3.7.3. Enregistrement dans le cadre d'une mesure de surveillance

Si le message est enregistré dans le cadre d'une mesure officielle de surveillance téléphonique au sens de l'art. 179^{octies} CP¹⁰, sa durée n'est pas limitée et le cumul des enregistrements peut atteindre des centaines d'heures. La sélection des échantillons est effectuée en

⁹ *infra* : Annexe II. Extraits du Code pénal suisse

¹⁰ *infra* : Annexe II. Extraits du Code pénal suisse

fonction du rapport de leur contenu à l'infraction pour laquelle la surveillance a été octroyée. La taille des champs lexicaux utilisés est en général étendue car le discours est construit, mais les thèmes se rapportent aux domaines des infractions considérées. Certaines organisations ou groupements évitent tout vocabulaire compromettant en utilisant des codes internes [NATARAJAN ET AL., 1995].

2.3.8. Influence du locuteur

Dans le domaine commercial, l'utilisation d'un système de contrôle d'accès est régulière et fréquente, la langue utilisée est déterminée et le locuteur désire être reconnu. Il coopère, d'une part en contrôlant son énoncé de manière à diminuer la variabilité intralocuteur et d'autre part en s'exprimant dans la langue demandée.

Dans le domaine forensique par contre, le locuteur n'est pas forcément coopératif et la langue parlée dépend uniquement des connaissances de chacun des interlocuteurs. D'autre part, l'intervalle de temps entre l'enregistrement de l'échantillon de parole inconnue et l'enregistrement de comparaison peut être long, de plusieurs mois à plusieurs années, selon les résultats de l'enquête et l'échéance de la prescription de l'infraction poursuivie [HOLLIEN, 1995].

2.3.8.1. Enregistrement dans le cadre d'un abus de téléphone

Dans ce premier cas, la voix peut être modifiée de façon involontaire, par les conditions psychologiques particulières de stress et de peur que peut engendrer le fait de commettre une telle infraction. La consommation de substances psychotropes et de tabac, ainsi que l'état de santé du locuteur, ont aussi une influence sur sa voix [HOLLIEN, 1990 ; BRAUN, 1994 ; HOLLIEN ET MARTIN, 1996]. Le locuteur peut aussi procéder à une modification volontaire du fonctionnement de l'un des organes participant à la production de la parole [HOLLIEN ET AL. ; 1982 ; GFROERER, 1994 IN : MASTHOFF, 1996]. LOCARD propose « l'introduction de tampons de coton enveloppant le gros bout d'une plume de coq dans les narines » ou « l'utilisation d'un dispositif de filtrage mécanique comme le mouchoir sur le combiné du téléphone » comme moyens de déguisement. Il est aussi possible de recourir à des dispositifs plus modernes de filtrage et de brouillage, analogiques ou numériques [LOCARD, 1932 ; MASTHOFF, 1996].

Les modifications qu'un locuteur peut apporter à un énoncé dans un but de déguisement peuvent porter sur la voix : respiration, registre, mode de phonation ; sur la parole : articulation, intonation, accent, vitesse d'élocution, stress, modification du tractus vocal par un corps étranger ou encore sur le contenu : jargon, dialecte étranger [GFROERER, 1994 IN : MASTHOFF, 1996 ; KÜNZEL, 1994A]. Certaines caractéristiques sont cependant plus difficiles à modifier que d'autres comme le montre le tableau récapitulatif d'une étude concernant les stratégies de déguisement présentée par FÄHRMANN [FÄHRMANN, 1966A ; FÄHRMANN, 1966B] (Figure II.6.).

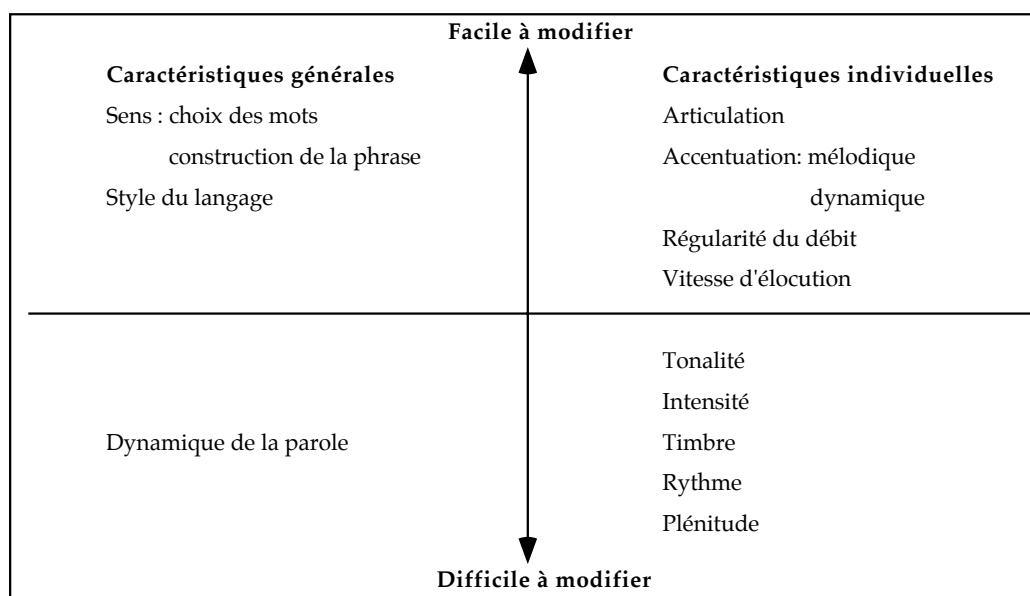


Figure II.6. Comparaison de la difficulté de modification de différentes caractéristiques de l'élocution [FÄHRMANN, 1966A]

Lorsque la stratégie de déguisement est laissée au libre choix du locuteur, on voit que, si la modification porte sur un seul paramètre, il s'agit d'un paramètre de la voix dans 30 % des cas et si elle porte sur plusieurs paramètres, la proportion qu'un paramètre de la voix soit modifié s'élève à 60 %. Dans 42 % des cas un ou plusieurs paramètres de la parole sont modifiés, mais le contenu n'est affecté que dans 22 % des cas. Dans 10 % des cas, des moyens de filtrage électroniques complètent les modifications intrinsèques [GFROERER, 1994 IN : MASTHOFF, 1996].

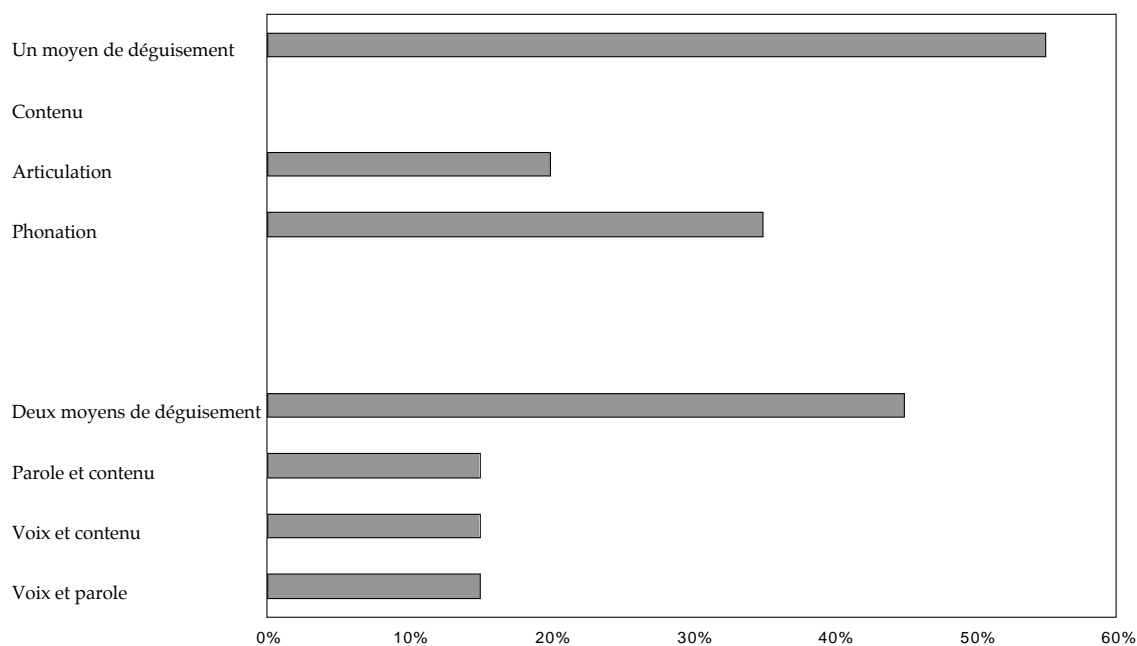


Figure II.7. Distribution des moyens de déguisements utilisés [MASTHOFF, 1996]

MASTHOFF confirme ces résultats ; il montre que lorsque le contenu est court, une seule phrase, le locuteur opte pour la modification d'un seul paramètre dans 55 % des cas. Le mode de phonation, paramètre lié à la voix, est modifié dans 30 % des cas et l'articulation, paramètre lié à la parole, dans 25 % des cas ; le contenu n'est jamais modifié seul. Dans 45 % des cas, deux déguisements sont utilisés, répartis en trois groupes de 15 % : phonation et articulation, phonation et contenu et articulation et contenu [MASTHOFF, 1996] (Figures II.6. et II.7., tableau II.2).

Moyen de déguisement	%	Particularités
Détails de la modification de la phonation		
Murmure	38	
Fréquence fondamentale rehaussée	31	Tous des hommes
Fréquence fondamentale abaissée	23	Toutes des femmes
Inspiration	8	Intelligible
Détails de la modification de l'articulation		
Imitation d'un dialecte	20	
Imitation d'un accent étranger	10	
Modification du tractus vocal	40	
Immobilisation de la langue	20	
Simulation d'une pathologie	10	
Détails de la modification du contenu		
Intonation non grammaticale	50	Modification du niveau d'intonation
Pauses non grammaticales	33	Pas de pauses ou pauses longues
Durées non grammaticales	17	Durée extrême des voyelles

Tableau II.2. Détails des moyens de déguisement [MASTHOFF, 1996]

Le résultat d'une étude de MCCLELLAND montre par contre que, lorsque le texte est long, le déguisement du contenu est la forme de déguisement préférée par les locuteurs. Malgré une liberté de choix totale, les locuteurs exploitent une douzaine de moyens de déguisement qu'ils combinent dans un peu moins de la moitié des cas, pour élaborer leur stratégie [MCCLELLAND, 1994 *IN* : MASTHOFF, 1996].

Selon [KÜNZEL, 1994A], il est clairement possible de déterminer qu'il y a déguisement dans environ 15 % des cas traités par le Bundeskriminalamt de Wiesbaden (BKA). REICH ET DUKE montrent qu'à partir d'un enregistrement, 89 % des auditeurs inexperts et 93 % des auditeurs experts détectent un déguisement librement choisi par des locuteurs¹¹ [REICH ET DUKE, 1979]. Ceux-ci admettent toutefois que les instructions données permettent aux locuteurs de choisir des réalisations extrêmes qui ne sont pas naturelles.

¹¹ *infra* : 4.3.3.3.2. Influence d'une modification de la voix

Pourtant, « tout déguisement n'implique pas comme résultat une élocution en dehors des variations produites par la population 'normale'. Le locuteur a un grand nombre de moyens d'altérer son élocution sans pour autant imiter consciemment un autre locuteur ou outrepasser les limites de la variation 'normale' de la population... Dès lors, aucune base ne laisse à penser, *a priori*, que les auditeurs sont capables d'évaluer de façon fiable, si l'élocution d'un locuteur est habituelle ou non pour un énoncé donné, mais, à un certain degré, chacune de ces modifications est susceptible de gêner l'identification de locuteurs » [NOLAN, 1983].

2.3.8.2. Enregistrement dans le cadre d'une mesure de surveillance

Dans ce second cas, l'ignorance d'être enregistré dans laquelle se trouve le locuteur n'induit vraisemblablement pas les modifications volontaires ou involontaires de la voix décrites *supra*. Par contre, la spontanéité qui découle de cette ignorance laisse au locuteur une substantielle faculté d'adaptation de son discours au contexte, à son humeur et aux différentes relations interpersonnelles qu'il entretient avec son interlocuteur [NOLAN, 1983 ; BROEDERS, 1995].

Dans un contexte formel, le locuteur tend à modifier les variables sociolinguistiques de son discours dans le sens de celles de personnes d'un statut supérieur [LABOV, 1972]. Un autre aspect du contexte est la notion sociologique d'étiquetage, qui dépend des rôles définis dans l'échange et de l'interprétation par le locuteur de son statut relatif dans cette interaction ; cet étiquetage va amener le locuteur à rapprocher ou à éloigner la forme de son discours de celle de son interlocuteur [GILES ET AL., 1979]. Une large variété d'informations peut aussi être introduite par le locuteur dans le discours, dans le but de présenter une personnalité correspondant à l'image qu'il veut montrer de lui-même [ARGYLE, 1976].

Ces informations sociolinguistiques ne sont pas invariantes, mais dépendent de l'interprétation que fait le locuteur des aspects sociaux de l'interaction [BROWN ET LEVINSON, 1979]. Le caractère formel des rapports que le citoyen peut entretenir avec l'autorité policière ou judiciaire, les rôles définis intrinsèquement liés à ces relations et l'image de soi-même qui peut vouloir être présentée dans de telles circonstances, permettent d'estimer l'étendue des différences entre une conversation se déroulant dans le cadre d'un interrogatoire ou d'une séance d'enregistrement de comparaison et une conversation téléphonique privée.

2.4. Conclusion

La plupart des éléments qui influencent négativement la qualité finale de l'indice matériel enregistré, comme la réponse en fréquence du transducteur ou l'état psychologique du locuteur, ne peuvent être améliorés dans le cadre de la procédure de collecte de cet indice. L'effort doit donc être concentré autour du seul élément sur lequel il est possible d'influer, le système d'enregistrement. Or, comme le relevait déjà BOLT en 1979, c'est malheureusement souvent le maillon le plus faible [BOLT, 1979]. Les forces de police de Suisse sont encore largement équipées de matériel d'enregistrement analogique obsolète, mais le moment est judicieux pour un passage à une stratégie d'enregistrement numérique, avec comme base des standards de qualité et des protocoles d'acquisition communs dans tout le pays.

L'avènement de l'ère de la transmission numérique de l'information sur les réseaux de télécommunication offre par contre la possibilité pour l'utilisateur final de recourir au cryptage de l'information transmise avec des algorithmes de cryptage puissants (*strong encryption*), inviolables par la seule puissance de calcul des ordinateurs actuels en un temps raisonnable. L'encryptage puissant des données informatiques est possible avec des algorithmes de cryptage de type *Data Encryption System* (DES) dont certains, comme *Pretty Good Privacy* (PGP), sont disponibles gratuitement sur le réseau Internet. Le cryptage des données audio est encore réservé à une certaine catégorie d'utilisateurs, mais des solutions commerciales, comme le concept de *Total Information Security*[®] développé par l'entreprise suisse *Crypto AG*, offrent des possibilités d'encryptage puissant non seulement aux armées, à la diplomatie et aux polices, mais aussi aux entreprises et aux personnes privées, pour leurs communications vocales et informatiques sur les différents réseaux téléphoniques, RTPC, RNIS, GSM et satellitaires.

Ce type de technologie est considéré comme sensible dans certains pays, dont la France et les Etats-Unis, et sa mise à disposition de tout utilisateur sans contrôle risquerait à moyen terme de rendre inopérante toute mesure officielle de surveillance des télécommunications. Dans un exposé intitulé « *Impact of Encryption on Law Enforcement and Public Safety* », présenté le 25 juillet 1996 devant la commission sur le commerce, les sciences et le transport du Sénat des États-Unis, le directeur du *Federal Bureau of Investigation* (FBI) constate que la dissémination sans contrôle de ce type de technologie d'encryptage peut être préjudiciable au travail des autorités chargées de l'application des lois et représenter à terme un risque majeur pour la sécurité publique [FREEH, 1996].

Des pays comme la France ont pris des mesures pour limiter l'utilisation de ce type d'algorithmes, par l'intermédiaire d'autorités de surveillance ; en Suisse par contre, ce type de problème semble absent du débat politique et des préoccupations des autorités chargées de faire respecter les lois.

III. METHODOLOGIE

3.1. Introduction

Le développement d'une méthode de reconnaissance automatique de locuteurs en vue d'une utilisation forensique implique des choix méthodologiques guidés par les exigences légales en matière de preuve scientifique, une connaissance des différentes méthodes de reconnaissance de locuteurs et des moyens de les évaluer. Il nécessite aussi l'adoption d'un processus d'inférence de l'identité du locuteur qui respecte à la fois la logique, l'approche scientifique et le rôle de l'expert dans le cadre du procès pénal.

3.2. Rôle de l'expert ou du scientifique

Pour éviter les pièges que les limites de la science mettent sous leurs pas, les scientifiques peuvent choisir trois attitudes [AIGRIN, 1996].

3.2.1. Le refus de témoigner

Quoique la question relève de leur domaine de compétence, certains scientifiques refusent de témoigner [AIGRIN, 1996]. Cette attitude de réserve est adoptée par le Bureau du Groupe de la Communication Parlée de la Société Française d'Acoustique (GFCP), de même que par la *British Association of Academic Phoneticians* (BAAP) [GFCP, 1991 ; BOË, 1998 ; BRAUN, 1995]. Ces groupes justifient leur position par le fait qu'à l'heure actuelle aucune méthode proposée n'est valide et, delà, personne n'est compétent pour donner une réponse fiable à la demande du monde judiciaire. C'est oublier la réalité du problème. Le refus de collaboration des hommes de l'art conduit l'ordre judiciaire, par manque de connaissance du domaine, à s'en remettre à l'expertise de personnes dont la compétence est de loin inférieure à celle des phonéticiens et des experts en science de la parole [BALDWIN ET FRENCH, 1990 ; BRAUN, 1995].

Cette situation est analogue à celle du grave problème qui existe dans le domaine de l'expertise judiciaire d'écriture manuscrite, notamment en France et, dans une moindre mesure, en Suisse [MATHYER, 1990 ; DAoust, 1995]. AIGRIN souligne d'autre part que cette « politique de l'autruche » ne lui semble pas digne de l'éthique scientifique [AIGRIN, 1996].

3.2.2. Le maximalisme

Cette attitude consiste à surévaluer, plus ou moins involontairement, les risques. Le phénomène n'est pas nouveau. Au XIX^e siècle déjà, Arago, opposé au développement des chemins de fer, présenta de manière apocalyptique les risques encourus, selon lui, par les usagers de ce nouveau mode de transport. Cette attitude permet à l'expert de se dédouaner si quelque chose tourne mal, mais elle ne résout pas les problèmes réels auxquels sont confrontés les décideurs et peut les conduire à des décisions injustifiées, parfois très coûteuses [AIGRIN, 1996].

Par exemple, l'absence d'étude, à grande échelle, de la variabilité de la plupart des caractéristiques dépendantes du locuteur observées et mesurées dans l'approche phonétique acoustique, force les experts à recourir à des probabilités subjectives pour interpréter leurs résultats. Or, il n'est pas possible de savoir dans quelle mesure leur expérience permet de combler les différences entre leurs probabilités subjectives personnelles et les probabilités statistiques, qui rendent compte de la variabilité réelle de ces caractéristiques ¹².

Ces incertitudes, de même que l'éthique qui anime certainement la plupart des experts, les poussent à maximiser les probabilités favorables à l'accusé et à minimiser celles qui lui sont défavorables, appliquant plus ou moins consciemment l'adage « *in dubio, pro reo* », alors qu'il s'agit là clairement du rôle du juge [MERMINOD, 1992].

3.2.3. La présentation et l'évaluation de l'état de l'art

Selon AIGRIN, la présentation des méthodes représentant l'état de l'art et l'évaluation de leurs capacités est la seule attitude vraiment scientifique, même si cette position peut demander beaucoup de travail et de sens critique, malgré le fait que la quête du risque zéro semble un leurre [AIGRIN, 1996].

3.2.4. Choix d'une approche méthodologique

Le présent travail s'efforce donc de présenter l'état de l'art dans le domaine de la reconnaissance de locuteurs en sciences forensiques, dans le but de déterminer l'habileté de chacune des trois approches actuellement pratiquées à inférer l'identité d'un locuteur. L'approche auditive, pratiquée par des profanes ou des experts et l'approche spectrographique font l'objet d'une étude bibliographique, alors que l'approche automatique fait l'objet d'une recherche théorique, bibliographique et expérimentale.

3.3. Exigences légales en matière de preuve scientifique

3.3.1. En droit suisse

En Suisse, la procédure pénale permet en principe de recevoir toutes les preuves. Les preuves obtenues par écoute téléphonique, enregistrement par magnétophone et enregistrement du numéro de téléphone de l'auteur d'appels répétés sont admises, sous certaines réserves. Si de telles preuves n'ont pas valeur d'aveu ou de preuve complète, elles peuvent néanmoins constituer des éléments susceptibles de s'ajouter à d'autres indices [PIQUEREZ, 1994].

Le juge apprécie librement la preuve qui lui est soumise, en faisant appel à son raisonnement. Cette libre appréciation n'est cependant pas illimitée et l'intime conviction ne dispense pas le magistrat d'utiliser une méthode logique dans l'évaluation des preuves qui lui sont présentées. Lorsque l'appréciation des preuves nécessite des connaissances particulières que le

¹² *supra* : 1.3. Le rôle des probabilités dans l'identification

juge ne possède pas, il est nécessaire que le magistrat ait recours à un homme de l'art ou un spécialiste, auquel il demande d'apporter sa collaboration à la manifestation de la vérité.

Cet expert judiciaire est nommé par l'instruction, au contraire de l'expert privé, nommé unilatéralement par l'une des parties au procès. Il se distingue du témoin, qui est appelé à relater ce qu'il a vu et entendu sans interpréter, alors que l'expert a pour mission d'éclairer le juge en donnant son appréciation technique sur certains points précis [PIQUEREZ, 1994].

3.3.2. En droit nord-américain

Dans le système juridique nord-américain par contre, la liberté de la preuve est limitée par le principe de recevabilité des preuves, dérivé du système ancien des preuves légales. L'interprétation de ce principe a permis le développement d'une réflexion avancée, notamment sur la recevabilité de la preuve scientifique en justice et de la méthodologie qui sous-tend la démonstration d'une telle preuve.

Selon la *Federal Rule of Evidence* (FRE) 901(b)(5)¹³, qui règle la recevabilité des preuves en matière d'identification, un témoignage reposant sur l'identification auditive d'un locuteur, entendu de manière directe ou par l'intermédiaire d'un système de transmission ou d'enregistrement, est recevable.

La preuve scientifique, quant à elle, est soumise à des règles de recevabilité depuis 1923, date à laquelle une cour Fédérale de justice des États-Unis a énoncé la première règle de recevabilité, connue sous le nom de standard de *Frye*¹⁴. La cour était confrontée à une preuve basée sur une théorie scientifique nouvelle et non familière, pour laquelle il n'existait aucun précédent ni principe établi permettant d'en déterminer la recevabilité. Elle conclut que la validité et la fiabilité de la technologie utilisée devaient être déterminées et décida que cette technologie devait recevoir l'acceptation générale des autorités « physiologiques et psychologiques », comme indication ou preuve de fiabilité ou de validité [LOEVINGER, 1995].

Le parti pris notoire des experts, choisis et payés par les parties, ainsi que les abus constatés dans le recours aux experts, ont conduit la Conférence Judiciaire à proposer de remplacer le standard de *Frye* par trois *Federal Rules of Evidence* (FRE), en 1961. Les *Federal Rules of Evidence* 701, 702 et 703¹⁵, qui ne sont devenues effectives qu'en 1975, autorisent l'expert à énoncer son témoignage basé sur une connaissance technique ou scientifique particulière, par exemple sous la forme d'une opinion. Elles permettent à ces témoignages de reposer sur des oui-dire ou sur des preuves non admissibles selon le standard de *Frye*, si elles sont reconnues par les scientifiques du domaine pertinent [LOEVINGER, 1995].

Ces règles générales n'étant pas satisfaisantes, la cour Suprême des États-Unis a décidé, en 1992, que la validité scientifique pour une application n'est pas forcément valable pour d'autres

¹³ *infra* : Annexe IV. Extraits des *Federal Rules of Evidence*

¹⁴ [Frye v United States (1923) 54 App DC 46, 293 F 1013, 34 ALR 145]

¹⁵ *infra* : Annexe IV. Extraits des *Federal Rules of Evidence*

applications du même domaine. Par conséquent, la détermination de la recevabilité d'une telle preuve requiert l'analyse des problèmes scientifiques dans les circonstances du cas, l'analyse des indices scientifiques de validité de la preuve et la détermination de la pertinence entre les problèmes et la preuve. Dans l'arrêt *Daubert v Merrel Dow Pharmaceuticals*¹⁶, et plus complètement dans l'arrêt *Conde v Velsicol Chemical Corporation*¹⁷, cette même cour énonce une série de règles concernant le témoignage scientifique et précise notamment des exigences concernant la méthodologie [LOEVINGER, 1995] :

Pour déterminer si un témoignage allégué, basé sur une connaissance scientifique, est valide et fiable du point de vue de la *Federal Rule of Evidence* 702¹⁸, six standards ont été établis par la cour. Les théories, le raisonnement ou la méthodologie, sur lesquels le témoignage repose, doivent : (1) être falsifiables, selon la définition de POPPER, avoir été testés ou pouvoir être testés ; (2) avoir fait l'objet d'une revue par les pairs (*peer review*) et de publication ; (3) avoir un taux d'erreur connu ou potentiel dans l'application ; (4) être généralement acceptés dans la communauté scientifique pertinente ; (5) être basés sur des faits ou des données dignes de confiance pour les experts du domaine ; (6) avoir une valeur probante qui n'est pas supplantée par les dangers d'un préjudice injuste, la confusion des conclusions ou l'induction en erreur du jury. Ces facteurs sont à considérer ensemble, aucun d'entre eux n'étant seul décisif ou déterminant [POPPER, 1973 ; LOEVINGER, 1995].

3.3.3. Choix d'une démarche

Une démarche scientifique n'est ni exigée ni privilégiée dans le système pénal suisse, mais une démarche respectant les critères de recevabilité des preuves scientifiques dans le système pénal nord-américain est choisie pour cette recherche. Ce choix est motivé par la volonté de limiter la part du raisonnement inductif au profit du raisonnement déductif dans le processus d'individualisation, et de développer une connaissance dans la ligne de l'enseignement prodigué et de la recherche développée à l'Institut de police scientifique et de criminologie de l'Université de Lausanne.

3.4. Méthodes de reconnaissance de locuteurs

3.4.1. Définitions

[HECKER, 1971] définit la reconnaissance de locuteurs comme : « tout processus de décision qui utilise des caractéristiques dépendantes du locuteur dans le signal de parole », alors que [ATAL, 1976] offre la formulation suivante : « tout processus de décision qui utilise quelques

¹⁶ [Daubert v Merrel Dow Pharmaceuticals (1993, US) 125 L Ed 2d 469, 113 S Ct 2786]

¹⁷ [Conde v Velsicol Chemical Corporation (1994) WL 184966, 6th Cir 1994]

¹⁸ *infra* : Annexe IV. Extraits des *Federal Rules of Evidence*

caractéristiques du signal de parole pour déterminer si une personne particulière est auteur d'un énoncé donné ».

Cette seconde formulation est préférable, car les processus de décision, requis dans le décodage du contenu linguistique d'un énoncé, font aussi appel à des caractéristiques du signal de parole dépendantes du locuteur. Par exemple, en reconnaissant une voyelle d'un locuteur inconnu, un auditeur peut interpréter des caractéristiques de ce locuteur inférées par le signal, comme son sexe, sur la base de la hauteur de la fréquence fondamentale [NOLAN, 1983].

3.4.2. Procédure

3.4.2.1. Extraction des caractéristiques

Comme aucune caractéristique spécifique au locuteur n'est actuellement identifiée dans la parole, son analyse présuppose une connaissance des aspects du signal de parole, en vue d'extraire les caractéristiques qui renferment les paramètres dépendant le plus manifestement de l'identité du locuteur. Pour être idéale, la caractéristique extraite devrait satisfaire aux critères suivants :

L'abondance : la caractéristique doit apparaître fréquemment dans la parole et ne pas engendrer de contraintes pour le locuteur.

L'efficacité : l'efficacité d'un paramètre pour une distinction des locuteurs est conditionnée par le rapport de sa variabilité intralocuteur à sa variabilité interlocuteur.

La mesurabilité : la caractéristique doit pouvoir être extraite dans un temps-microprocesseur raisonnable, même si une extraction en temps réel n'est pas une priorité pour les applications forensiques.

L'infailibilité : la caractéristique et sa distribution ne doivent pas pouvoir être modifiées par un effort conscient du locuteur. Elle doit être telle qu'un imposteur ne puisse réussir une tentative d'imitation.

La pérennité : la caractéristique et sa distribution doivent rester stables au cours du temps pour un locuteur donné ; en particulier elle ne doit pas être affectée par des éléments perturbant le locuteur tels que la santé, l'état émotionnel ou le contexte de la communication.

La robustesse : la caractéristique doit être insensible aux perturbations du signal de parole occasionnées par la prise de son, le canal de transmission ou le système d'enregistrement. [WOLF, 1972 ; BOITE ET KUNT, 1987 ; THEVENAZ, 1993 ; HOMAYOUNPOUR ET CHOLLET, 1995]

L'analyse de la sélection des caractéristiques, développée par KWAN dans le domaine forensique, concorde avec ces critères, mais l'auteur relève avec pertinence l'importance d'une distribution rectangulaire des caractéristiques et de l'absence de corrélation entre les caractéristiques analysées, en vue d'augmenter le pouvoir discriminatoire de la méthode [KWAN, 1977]. BREMERMAN observe d'autre part que le choix judicieux des caractéristiques conditionne plus de la moitié de l'efficacité de l'identification ; aucun traitement mathématique postérieur ne saurait combler des caractéristiques mal choisies [BREMERMAN, 1971 IN : KWAN, 1977].

Selon la méthode adoptée, l'extraction des caractéristiques dépendantes du locuteur est opérée soit par un expert, de manière auditive et à l'aide de moyens de visualisation, soit automatiquement, sur la base d'algorithmes d'analyse du signal de parole. Les caractéristiques prises en compte par les experts sont généralement liées à la réalité physiologique, comme la fréquence fondamentale ou la hauteur et l'étendue des formants, alors que les caractéristiques extraites automatiquement sont plutôt liées à la réalité du traitement numérique des signaux ; il peut s'agir de coefficients de prédiction linéaire, de coefficients spectraux ou cepstraux, ou encore de vecteurs obtenus par quantification vectorielle¹⁹. Malheureusement aucune caractéristique satisfaisant à tous les critères énoncés *supra* n'a encore été isolée dans le signal de parole.

3.4.2.2. Comparaison des caractéristiques

La comparaison des caractéristiques extraites est réalisée soit de manière subjective par un expert, qui évalue les ressemblances et les différences entre les caractéristiques extraites de la voix inconnue et celles extraites de la voix de la personne mise en cause, présentes dans l'enregistrement de comparaison. Cette comparaison peut aussi être effectuée de manière objective, à l'aide d'une méthode de reconnaissance automatique de locuteurs, qui fournit un estimateur numérique, représentant une distance mathématique ou une proximité statistique entre la voix inconnue et la voix présente sur l'enregistrement de comparaison. Le résultat de cette évaluation représente l'élément de preuve, qui peut exprimer soit une probabilité de coïncidence fortuite, soit une opinion subjective de la fréquence des caractéristiques analysées.

3.4.3. Classification des méthodes de reconnaissance

3.4.3.1. Selon le type de méthode

3.4.3.1.1. Définition

Sous le titre général de reconnaissance de locuteurs, il est nécessaire de distinguer un certain nombre de domaines d'étude distincts. HECKER reconnaît trois divisions majeures : la reconnaissance de locuteurs par audition, par machine et par comparaison visuelle de spectrogrammes [HECKER, 1971].

La reconnaissance de locuteurs par audition, *Speaker Recognition by Listening* (SRL), est constituée par l'étude de la manière dont les auditeurs humains réalisent la tâche d'association d'une voix particulière à un individu particulier ou à un groupe et notamment dans quelles circonstances une telle tâche peut être remplie [NOLAN, 1983].

La reconnaissance de locuteurs par machine, *Speaker Recognition by Machine* (SRM) [HECKER, 1971], ou reconnaissance automatique de locuteurs, *Automatic Speaker Recognition* (ASR) [O'SHAUGNESSY, 1986], est l'étude de la capacité de l'outil informatique à procéder à la tâche de reconnaissance de locuteurs, sur la base de méthodes exploitant la théorie de l'information, la reconnaissance automatique de formes et l'intelligence artificielle perceptive [BUNGE, 1991]. Elle est souvent

¹⁹ *infra* : 6.2.3. Approches actuelles

considérée comme objective par rapport à la SRL, à cause de sa relative indépendance de la décision humaine subjective [NOLAN, 1983].

La reconnaissance de locuteurs par spectrogrammes, *Speaker Recognition by Spectrograms (SRS)*, consiste en la prise de décision sur l'identité ou la non-identité du locuteur sur la base de la comparaison visuelle de spectrogrammes vocaux par des observateurs entraînés. Ces spectrogrammes sont aussi appelés vocogrammes [KERSTA, 1973] ou sonagrammes [NOLAN, 1983].

3.4.3.1.2. Aspects forensiques

Les trois approches sont actuellement pratiquées en sciences forensiques. La reconnaissance de locuteurs par audition est pratiquée soit par des experts, phonéticiens ou spécialistes des sciences de la parole, soit par des profanes, principalement les victimes ou les témoins d'une infraction. La reconnaissance de locuteurs par spectrogrammes, en tant que méthode à part entière, est surtout pratiquée aux États-Unis par des examinateurs de spectrogrammes, alors que la reconnaissance de locuteurs par machine commence à être utilisée sous forme de systèmes semi-automatiques ou de systèmes assistés par ordinateur [KÜNZEL, 1994A ; FALCONE ET DE SARIO, 1994].

3.4.3.2. Selon le type d'approche, subjective ou objective

3.4.3.2.1. Définition

Une autre classification, bipartite, a été proposée par [LEWIS, 1984] et [DODDINGTON, 1985]. Elle est basée sur le type d'approche, subjective ou objective, utilisé dans les deux étapes de la reconnaissance de locuteurs : l'extraction des caractéristiques et leur comparaison.

3.4.3.2.2. Aspects forensiques

Dans la reconnaissance de locuteurs par audition, qu'elle consiste en l'opinion d'un phonéticien, d'une personne familière de la personne mise en cause, d'un témoin ou d'une victime, l'extraction et la comparaison de ces caractéristiques sont effectuées de manière subjective. Dans la comparaison visuelle de spectrogrammes vocaux, l'extraction des caractéristiques est plus objective, car elle fait appel à un instrument, le spectrographe sonore, mais la comparaison demeure subjective [GRUBER & POZA, 1995].

Le but de la reconnaissance automatique, par contre, est de tendre vers une extraction et une comparaison objectives des caractéristiques en utilisant, pour la première, diverses techniques de traitement du signal et, pour la seconde, des systèmes de classification automatique [DEVIJVER ET KITTLER, 1982 ; LEWIS, 1984].

Cependant, dans chacune de ces approches, tout ou partie du processus de reconnaissance demeure subjectif. « Une procédure adéquate vise à réduire le facteur humain autant que possible, bien que le plus sophistiqué des systèmes nécessite l'interaction d'un expert à de nombreuses reprises, en commençant par la sélection d'énoncés de paroles adéquats, le filtrage et le prétraitement des signaux, la sélection ou la suppression de certains paramètres acoustiques ou linguistiques, l'interprétation des résultats chiffrés et la formulation de la décision finale d'identité ou de non-identité » [KÜNZEL, 1994A].

3.4.3.3. Selon le mode de dépendance au texte

3.4.3.3.1. Définition

Cette troisième classification distingue les méthodes indépendantes du texte des méthodes dépendantes du texte. Les premières offrent la possibilité d'utiliser des échantillons formés de n'importe quels mots ou phrases, alors que les secondes requièrent que l'échantillon de parole inconnue soit formé des mêmes mots ou phrases que l'échantillon de comparaison [FURUI, 1994]. Une méthode parfaitement indépendante du texte devrait aussi satisfaire à la condition d'indépendance par rapport à la langue parlée.

Trois niveaux de dépendance au texte peuvent cependant être distingués. La dépendance rigoureuse au texte, qui nécessite que les échantillons soient formés de la même séquence de mots, la dépendance au vocabulaire, qui permet l'utilisation d'échantillons formés de séquences de mots différents choisis à l'intérieur d'un vocabulaire restreint et la dépendance à l'événement de parole, qui ne nécessite que la présence de certains événements phonétiques particuliers dans les échantillons [BIMBOT ET AL., 1994].

3.4.3.3.2. Aspects forensiques

L'absence complète de maîtrise du criminaliste sur l'indice milite en faveur de l'utilisation de méthodes de reconnaissance indépendantes du texte. La reconnaissance de locuteurs par l'audition est indépendante du texte lorsqu'elle est effectuée par des profanes et ne nécessite qu'une dépendance à certains événements phonétiques particuliers, présents dans les différents échantillons, lorsqu'elle est réalisée par un expert. La reconnaissance de locuteurs par spectrogrammes, démarche essentiellement comparative, exige une dépendance plus rigoureuse par rapport au texte²⁰. La reconnaissance automatique de locuteurs connaît des méthodes dépendantes et indépendantes du texte.

3.4.4. Choix d'une méthode

Dans cette recherche, plusieurs raisons ont conduit à considérer l'approche automatique d'un point de vue théorique et expérimental, plutôt que les approches auditive ou spectrographique. Premièrement, la démarche scientifique choisie tend vers une minimisation du facteur humain et vers une objectivation des méthodes d'analyse. Deuxièmement, seule une procédure automatisée permet d'aborder le problème d'une manière réellement statistique, par la prise en compte d'un nombre significatif d'hypothèses alternatives²¹. Finalement, l'automatisation s'inscrit dans la tendance actuelle d'optimisation des processus inclus dans le domaine de l'expertise forensique. Dans son rapport « *Voice Analysis* » présenté au congrès de l'Interpol en 1998, BRAUN souligne d'ailleurs que la recherche tend actuellement à se focaliser sur des procédures plus objectives et moins gourmandes en temps de travail [BRAUN, 1998].

²⁰ *infra* : 5.4.3. Les standards de l'IAI

²¹ *infra* : 3.5.5. Choix d'un processus d'inférence de l'identité

Une procédure d'évaluation de l'approche phonétique, bien qu'absolument nécessaire, reste extrêmement difficile à mettre sur pied, pour plusieurs raisons : le faible nombre de personnes pratiquant l'expertise en reconnaissance de locuteurs dans la même langue, le temps nécessaire à l'analyse d'un seul cas, la constitution de cas fictifs de difficulté comparable dans des langues différentes, l'absence de consensus et d'unité autour des procédures d'analyse²² et une absence de volonté de transparence parmi les principaux intéressés. Il est en effet frappant de constater la rareté des publications concernant l'explicitation de la méthodologie dans le domaine de l'expertise phonétique [KÜNZEL, 1987 ; KOVAL ET AL., 1998A ; KOVAL ET AL., 1998B].

Ces éléments expliquent, en partie, l'absence d'évaluation à grande échelle et de contrôle de qualité dans la phonétique forensique, alors qu'un système de *proficiency testing* existe dans beaucoup de domaines des sciences forensiques, y compris dans un domaine analogue, celui de l'expertise d'écriture manuscrite.

L'évaluation des examinateurs de spectrogrammes est encore plus difficile à réaliser depuis l'Europe, puisque cette pratique existe surtout aux États-Unis. Par contre, les nombreuses études de l'approche spectrographique réalisées à ce jour permettent d'estimer la validité et la fiabilité de cette approche.

3.5. Inférence de l'identité d'un locuteur

L'inférence de l'identité d'un locuteur à partir de l'élément de preuve issu de l'analyse de la voix a été envisagée de plusieurs manières dans la littérature. Les auteurs ont principalement été influencés par les processus de décision utilisés dans les autres domaines des sciences forensiques et dans les autres applications de la reconnaissance automatique de locuteurs. Pour certains auteurs, il s'agit d'un problème de discrimination alors que pour d'autres il s'agit d'un problème de classification. D'autres encore estiment que les experts devraient se contenter de quantifier les taux d'erreur de type I et II²³ des méthodes qu'ils utilisent. Seul un petit nombre pense que la solution passe par l'évaluation de rapports de vraisemblance.

3.5.1. Discrimination

3.5.1.1. Définition

La discrimination, aussi appelée vérification de locuteurs dans le domaine de la reconnaissance automatique de locuteurs, consiste à mesurer une distance entre un enregistrement de parole inconnue et un enregistrement de comparaison et à prendre une décision binaire d'acceptation ou de rejet en comparant cette distance à un seuil établi *a priori* (Figure III.1.) [O'SHAUGNESSY, 1986]. THEVENAZ nomme cette tâche « vérification de locuteurs par acceptation », qu'il distingue de la « vérification de locuteurs par rejet », où l'enregistrement de parole inconnue est comparé à tous les enregistrements de comparaison connus [THEVENAZ, 1990]. Tous les

²² *infra* : 4.4. Procédure de reconnaissance par des experts

²³ *infra* : 3.5.3. Quantification des taux d'erreur de type I et de type II

enregistrements de comparaison doivent en principe être rejetés, à l'exception de celui dont le locuteur revendique l'identité. La décision finale, combinaison des décisions partielles, est binaire : acceptation ou rejet.

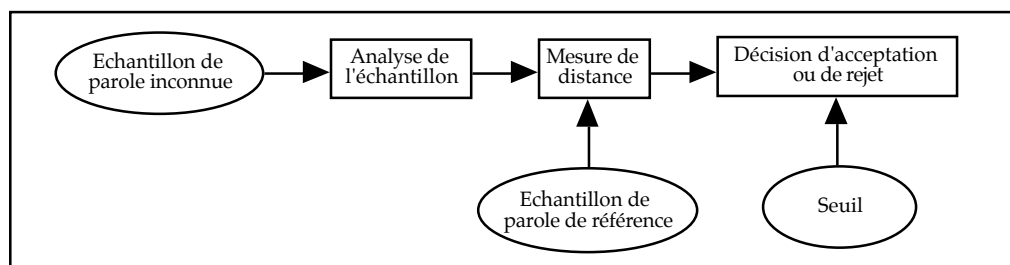


Figure III.1. Structure du système de vérification de locuteurs [FURUI, 1994]

La discrimination amène deux types d'erreurs : le faux rejet, faux négatif ou erreur de type I, lorsque le locuteur authentique n'est pas accepté comme tel, et la fausse acceptation, faux positif ou erreur de type II, lorsqu'un imposteur est pris pour le locuteur qu'il prétend être [TOSI, 1981, DODDINGTON, 1985, BIMBOT *ET AL.*, 1994]. La variation du seuil permet d'optimiser le taux de faux rejet ou de fausse acceptation.

3.5.1.2. Analyse forensique

DODDINGTON a proposé la vérification de locuteurs comme processus d'inférence de l'identité du locuteur en sciences forensiques, mais il précise que les décisions devraient être basées sur un modèle statistique valide ; il reconnaît aussi qu'un modèle efficace est difficile à établir dans cet environnement, à cause du manque de contrôle sur le signal de parole et de la difficulté d'appréciation des conditions acoustiques et de transmission [DODDINGTON, 1985]. Du même point de vue, NOLAN souligne aussi que le processus d'inférence de l'identité est plus proche de la discrimination que de la classification, puisque deux enregistrements sont comparés et qu'un seuil de similarité est appliqué implicitement ou explicitement [NOLAN, 1990].

La décision de discrimination entre l'indice matériel et la personne mise en cause dépend d'un seuil, qui peut être qualitatif : une opinion subjective au sujet des ressemblances et des différences, ou quantitative : un indice numérique de proximité. Ce point de vue conduit à considérer la discrimination comme une exclusion et la non-discrimination comme une identification. Or ce concept de l'identité ne correspond pas à la définition de l'individualisation forensique. Si la probabilité de coïncidence fortuite n'est pas nulle, corollaire du seuil, la conclusion d'identification est inadéquate et erronée [CHAMPOD ET MEUWLY, 1998].

De plus le seuil est par essence une qualification du niveau acceptable de doute raisonnable adopté par l'expert. Par contre, les juristes interprètent ce seuil comme une identification de locuteurs « au delà du doute raisonnable » [CHAMPOD ET MEUWLY, 1998]. Les juristes accepteraient-ils que le concept de doute raisonnable échappe à leur prérogative et que le seuil soit imposé à la cour par le scientifique ? Dans la doctrine, la réponse est négative, comme l'a exprimé le *Panel on Statistical Assessments as Evidence in Courts* :

« ... en ce qui concerne les éléments de preuve, la loi peut établir des seuils différents de ceux que les statisticiens considèrent comme suffisants pour conclure. Clairement, la loi doit prévaloir et le statisticien doit s'ajuster sur les standards légaux. Autrement dit, c'est la fonction d'utilité de la cour qui est appropriée, et non celle du statisticien » [FIENBERG, 1989 IN : CHAMPOD ET MEUWLY, 1998].

Pour cette raison la vérification de locuteur est inadaptée à l'inférence de l'identité du locuteur en sciences forensiques. Suite à un sondage parmi les phonéticiens sur l'utilisation des échelles de probabilité, NOLAN a proposé l'adoption de la démarche d'évaluation de rapports de vraisemblance, sur la base d'une présentation détaillée de « *Interpreting Evidence: Evaluating Forensic Science in the Courtroom* » de ROBERTSON ET VIGNAUX, lors de la conférence annuelle de l'IAFP à Edinbourg en 1997 [ROBERTSON ET VIGNAUX, 1995]. Ce point de vue est parfaitement en accord avec le poster intitulé « *Likelihood ratios for automatic speaker recognition in forensic applications* », présenté par MEUWLY ET DRYGAJLO lors de la même conférence [MEUWLY ET DRYGAJLO, 1997]. Lors de la discussion qui a suivi la présentation de CHAMPOD à Avignon en 1998, DODDINGTON a aussi reconsidéré son point de vue et admis que l'évaluation de rapports de vraisemblance est la démarche la plus adéquate de l'inférence de l'identité du locuteur en sciences forensiques [CHAMPOD ET MEUWLY, 1998].

3.5.2. Classification

3.5.2.1. Définition

La classification, aussi appelée identification de locuteurs dans le domaine de la reconnaissance automatique de locuteurs, consiste à comparer un enregistrement de parole inconnue à chacun des enregistrements de comparaison présents dans un ensemble de référence et d'établir un classement des enregistrements de comparaison [DODDINGTON, 1985, O'SHAUGNESSY, 1986].

3.5.2.1.1. Classification en ensemble fermé (*closed set*)

Lorsqu'il est possible de déterminer *a priori* qu'il existe un enregistrement de comparaison appartenant au locuteur testé dans l'ensemble de référence, celui-ci peut être considéré comme un ensemble fini de N échantillons de référence et la décision à prendre est en 1 sur N (Figure III.2.) [CORSI, 1982 ; FURUI, 1994].

Cette tâche de classification ne peut être entachée que d'un seul type d'erreur : la fausse identification. Elle apparaît lorsque l'échantillon de référence, dont la distance à l'échantillon de parole inconnue est minimale, ne correspond pas à l'échantillon de référence du locuteur authentique. La probabilité d'erreur augmente avec la taille de l'ensemble de référence et à chaque comparaison est associée une probabilité d'erreur finie et non nulle [CORSI, 1982, BIMBOT ET AL., 1994].

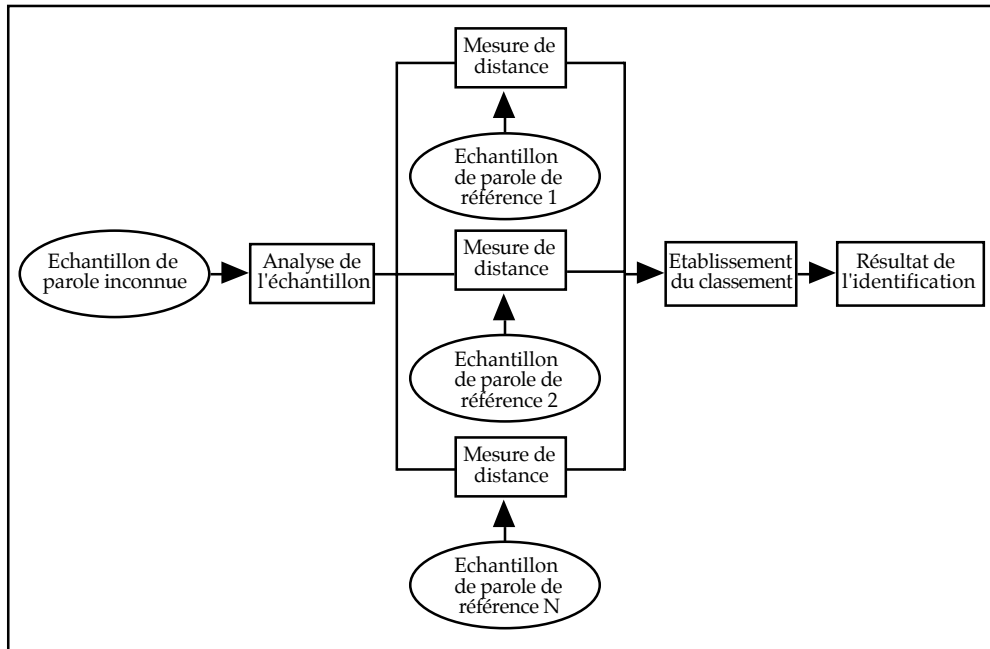


Figure III.2. Structure du système d'identification en ensemble fermé [FURUI, 1994].

3.5.2.1.2. Classification en ensemble ouvert (*open set*)

Lorsqu'il n'est pas possible de déterminer *a priori* s'il existe un enregistrement de comparaison appartenant au locuteur testé dans l'ensemble de référence, celui-ci doit être considéré comme un ensemble ouvert de $N+1$ échantillons. La décision devient binaire : acceptation ou rejet, car le résultat de la classification est soumis à un seuil établi *a priori*. La structure de ce système correspond à la juxtaposition des deux précédents (Figure III.3.) [BIMBOT ET AL., 1994].

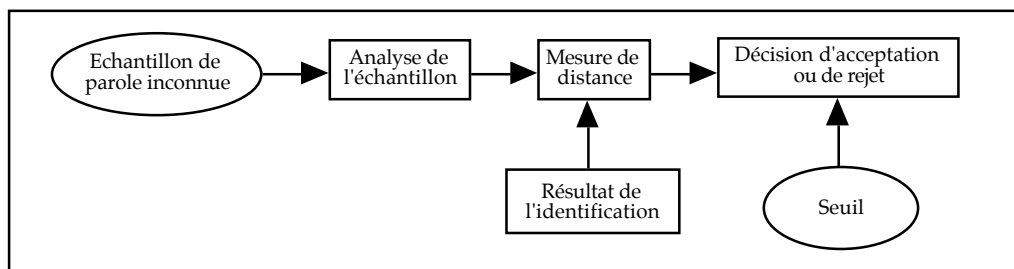


Figure III.3. Structure du système d'identification en ensemble ouvert

Dans une classification en ensemble ouvert, trois types d'erreur sont à prendre en compte : le faux rejet qui conduit à une fausse élimination et deux types de fausse acceptation qui conduisent à une fausse identification : l'échantillon de parole inconnue est faussement mis en relation avec un enregistrement de comparaison de l'ensemble de référence alors que :

- a. un autre enregistrement de comparaison de l'ensemble de référence concorde ;
- b. aucun enregistrement de l'ensemble de référence ne concorde.

Dans sa définition du concept de population potentielle, KWAN reprend les qualificatifs de *closed set* et d'*open set*, sur la base des définitions données par BOLT dans le domaine de la reconnaissance de locuteurs [BOLT ET AL., 1969 ; KWAN, 1977]. Curieusement il utilise ces deux qualificatifs dans un autre sens que BOLT, en déterminant que les conditions définissant l'*open set* sont remplies lorsqu'il n'y a pas de connaissance *a priori* de caractéristiques génériques du locuteur comme la langue qu'il parle et que les conditions du *closed set* sont remplies lorsque des caractéristiques de ce type sont connues. Comme cette divergence de terminologie peut amener des confusions, la terminologie originale de BOLT est préférée, en attendant qu'une terminologie adéquate soit trouvée pour définir les qualifications décrites par KWAN, qui sont différentes.

3.5.2.2. Analyse forensique

KÜNZEL considère que l'inférence de l'identité du locuteur ne peut être assimilée à une discrimination car, selon lui, un système fonctionnant sur la base d'un seuil établi *a priori* n'est pas concevable. Comme la variation du seuil permet d'optimiser les taux de faux rejet ou de fausse acceptation, ces taux ne sont jamais simultanément nuls. Cette probabilité d'acquiescement d'un coupable ou de condamnation d'un innocent lui apparaît comme inconcevable du point de vue éthique. Cette question ne peut pas non plus être assimilée à une classification en ensemble fermé, puisqu'il n'est pas possible de déterminer *a priori* si le locuteur inconnu se trouve dans l'ensemble des locuteurs suspects. Il conclut donc que l'inférence de l'identité du locuteur en sciences forensiques doit être envisagée sous l'angle de la classification en ensemble ouvert, car l'ensemble des locuteurs potentiels est pratiquement illimité et ne peut être restreint que par des informations concernant la langue parlée et le sexe [KÜNZEL, 1994A].

En fait, la classification ne peut se concevoir dans un ensemble fermé de locuteurs, car la décision de l'exhaustivité des personnes mises en cause qui forment la population potentielle n'est pas du ressort de l'expert, mais appartient à la cour. De plus il semble particulièrement arbitraire de révéler seulement le résultat concernant le meilleur candidat, sans fournir celui obtenu par les autres, comme le montre notamment l'exemple de WALSH, dans le domaine de l'interprétation du verre. Dans un cas de cambriolage une fenêtre est brisée par l'auteur ; deux personnes sont mises en cause après que des petits fragments de verre ont été retrouvés sur leurs vêtements respectifs. Une analyse de l'indice de réfraction du verre montre une concordance des indices avec la fenêtre brisée. Dans un cadre d'interprétation faisant appel à une approche continue des données, la probabilité de coïncidence fortuite entre le verre de la vitrine et l'indice est estimée à 1/1100 pour la première personne mise en cause, et à 1/900 pour la seconde. Concevoir l'identification forensique comme une classification en ensemble fermé revient à déclarer la première personne comme identifiée et à focaliser injustement l'élément de preuve sur cette dernière. En effet, l'élément de preuve ne favorise pas de façon évidente l'hypothèse que l'auteur soit la première personne mise en cause, par rapport à l'hypothèse que l'auteur soit la seconde personne mise en cause [WALSH ET AL., 1996 IN : CHAMPOD ET MEUWLY, 1998].

Pour surmonter cette carence, la classification devrait être envisagée dans un ensemble ouvert de locuteurs, mais ce processus d'inférence de l'identité inclut une discrimination finale

basée sur un seuil, qui lui confère les mêmes inconvénients que la discrimination [CHAMPOD ET MEUWLY, 1998].

3.5.3. Quantification des taux d'erreur de type I et de type II

3.5.3.1. Définition

Pour mesurer les performances d'un processus de décision impliquant des humains et des instruments de mesure, les statisticiens ont souvent recours à des fonctions de coût des erreurs de faux rejet et de fausse acceptation, comme la *Receiver Operating Characteristic* (ROC) (Figure III.4). Cette méthode a été proposée par BOLT, pour évaluer les performances des examinateurs de spectrogrammes et par PAOLONI, pour celle des systèmes automatiques de vérification et d'identification *open set* utilisés en sciences forensiques [BOLT ET AL., 1979 ; PAOLONI ET AL., 1994].

Une autre évaluation courante consiste à calculer les taux d'erreur de type I et de type II du classificateur et de déterminer la valeur pour laquelle l'erreur de type I est égale à l'erreur de type II. Cette valeur est connue sous le nom de taux d'égalité d'erreur ou *equal error rate* (EER) [FURUI, 1997].

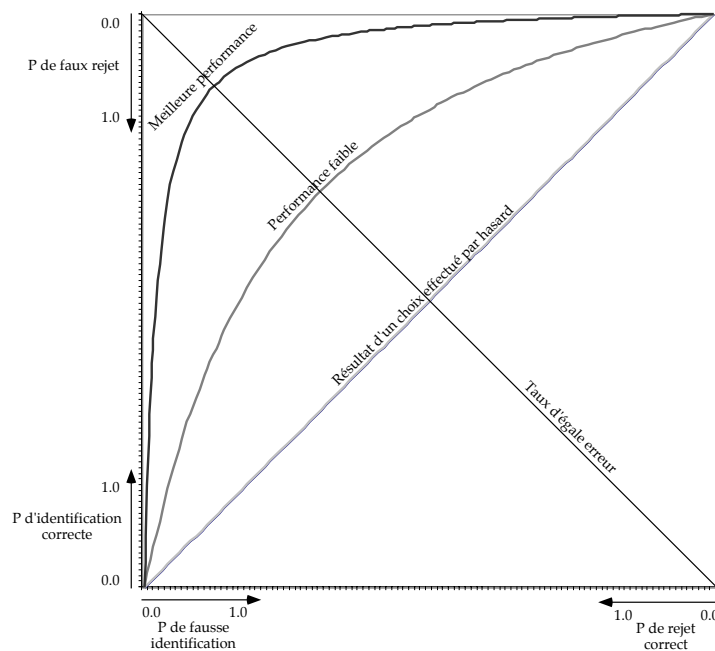


Figure III.4. Représentation graphique de trois courbes ROC et du taux d'égalité d'erreur, d'après [GRUBER ET POZA, 1995]

3.5.3.2. Analyse forensique

Certains auteurs pensent que, soit la discrimination soit la classification en ensemble ouvert sont envisageables, selon les circonstances du cas [BIMBOT ET AL., 1994 ; PAOLONI ET AL., 1994 ; ROSE ET DUNCAN, 1995]. Dans ces deux cas les décisions possibles peuvent être représentées dans un tableau, en fonction des différents états (Tableau III.1.).

		État	
		Identité (ID)	Non-identité (\bar{ID})
Décision	Identification (+)	0,99	0,01 (Erreur de type I)
	Non-identification (-)	0,01 (Erreur de type II)	0,99

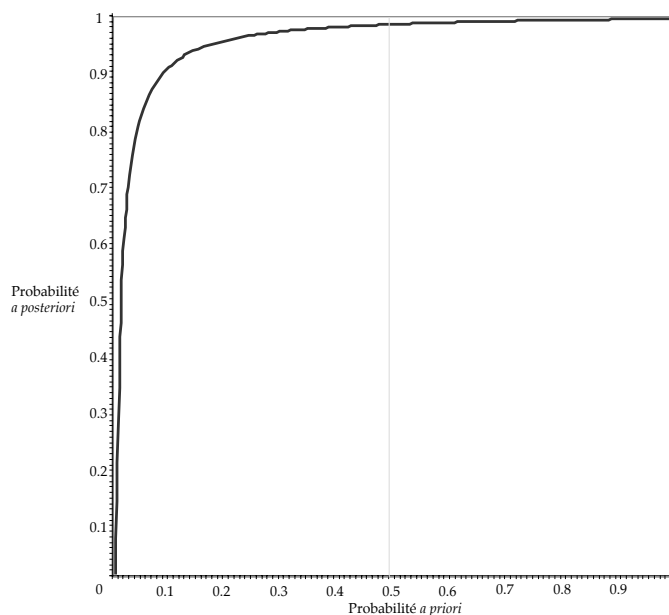
Tableau III.1. Représentation des taux d'erreur de type I et de type II pour la tâche de discrimination

Admettons qu'un indice matériel enregistré soit comparé à la voix d'une personne mise en cause et accepté par un système dont les performances correspondent à celles présentées dans le tableau III.1. Cette décision positive d'identification permet-elle à l'expert de conclure que la personne mise en cause est très probablement l'auteur de l'indice matériel ?

La reformulation de la conclusion en termes de probabilités indique que la personne est correctement identifiée dans 99% des cas, $P(ID|+) = 0,99$. Selon le théorème de Bayes, la probabilité *a posteriori* dépend partiellement du résultat de l'analyse, mais aussi de la probabilité *a priori* de l'identité, $P(ID)$:

$$P(ID|+) = \frac{P(+|ID) P(ID)}{P(+|ID) P(ID) + P(+|\bar{ID}) P(\bar{ID})} = \frac{0,99 P(ID)}{0,99 P(ID) + 0,01 (1 - P(ID))} \quad (3.1)$$

La validité de la conclusion dépend donc sérieusement de la probabilité *a priori* de l'identité. Le même type de démonstration a été présenté pour l'interprétation de la preuve par analyse d'ADN [BALDING ET DONNELLY, 1994 IN : CHAMPOD ET MEUWLY, 1998]. La représentation graphique proposée par BERRY (Figure III.5.) montre que $P(ID|+) \geq 0,99$ si, et seulement si $P(ID) \geq 0,5$ (Figure III.4.) [BERRY, 1991 IN : CHAMPOD ET MEUWLY, 1998].

Figure III.5. Évolution de la probabilité *a posteriori* en fonction de la probabilité *a priori*

Pour cette raison, il est faux de prétendre que l'identité du locuteur est démontrée avec un taux d'erreur de 1%, car cette conclusion implique que l'expert tienne pour vrai une probabilité *a priori* de 0,5. Dans le domaine de la recherche en paternité, cette pratique a même été qualifiée de « neutre », puisque si la personne suspectée est hors de cause, une seule autre personne est concernée [HUMMEL, 1984 IN : TARONI ET AITKEN, 1996]. Cette vision est cependant arbitraire, comme toute autre qui vise à la détermination de la valeur de la probabilité *a priori* par l'expert, car ce point relève de la compétence du juge et du jury²⁴ [EVETT, 1983 ; TARONI ET AITKEN, 1996].

3.5.4. Évaluation de rapports de vraisemblance

L'analyse des méthodes précédentes montre que l'inférence de l'identité d'un locuteur à partir d'un élément de preuve fourni par l'analyse de la voix ne peut être envisagée d'un point de vue déterministe par l'expert, en termes de culpabilité ou d'innocence. Le rôle du scientifique se borne à établir la vraisemblance de l'élément de preuve en cas d'identité ou de non-identité de la personne mise en cause.

3.5.4.1. Définition

Cette approche par évaluation des rapports de vraisemblance prend sa source dans le théorème de Bayes. Par rapport à la statistique classique, le point de vue bayésien se distingue notamment par la prise en considération des probabilités *a priori* des hypothèses vérifiées. Ainsi le niveau de signification, la probabilité du premier type d'erreur fixé habituellement à une valeur faible (cinq ou un pour cent), est en réalité dans l'interprétation bayésienne, fonction à la fois de la probabilité subjective qu'on attribue au premier type d'erreur et du coût de l'erreur qu'on accepte de courir [MATALON, 1967].

3.5.4.2. Exemple

« Un joueur se demande si son adversaire triche. Il ne dispose d'aucune preuve parfaitement convaincante, mais seulement d'un certain nombre d'indices : l'adversaire a l'air louche, il a gagné huit parties sur dix, etc. Aucun de ces indices n'est suffisant : un joueur honnête peut parfaitement « avoir de la chance » et gagner souvent. Quant à l'air louche, c'est une question d'appréciation. Supposons qu'on puisse évaluer la probabilité conditionnelle qu'un joueur gagne huit parties sur dix, s'il triche ; on peut raisonnablement penser qu'elle est assez élevée. Mais cela ne nous suffit pas pour l'accuser ; ce qu'il nous faudrait, pour être en mesure de tirer une conclusion inverse, c'est la probabilité conditionnelle inverse, la probabilité pour qu'un individu triche, sachant qu'il a gagné huit parties sur dix » [MATALON, 1967].

3.5.4.3. La notion de probabilité *a priori*

Le problème a été abordé sous cette forme à la fin du XVIII^e siècle par le Révérend Thomas Bayes (702 – 1752), qui a cherché à calculer la « probabilité des hypothèses », et le théorème qui porte son nom nous donne cette probabilité. Sa démonstration à partir des axiomes ne soulève aucune difficulté et son application n'exige que des calculs simples. Toutefois la formule présente une caractéristique essentielle : elle nous indique que si l'on veut être en mesure de calculer la

²⁴ *infra* : 3.5.4.4. Application forensique

probabilité pour qu'un joueur ait triché, sachant qu'il a gagné huit fois sur dix et qu'un tricheur gagne avec une probabilité de quatre-vingt-dix pour cent, il est indispensable en plus d'avoir une idée *a priori* sur la probabilité de tricher, avant de disposer d'aucun indice. En d'autres termes, le théorème de Bayes ne permet pas de se faire une opinion à partir des indices uniquement, mais indique simplement comment notre jugement préalable doit être modifié par le rapport de vraisemblance ou *Likelihood Ratio* (LR), calculé à partir de ces indices [MATALON, 1967].

3.5.4.4. Application forensique

3.5.4.4.1. Principe

L'application du théorème de Bayes au problème juridique pénal a été initiée par KAPLAN [KAPLAN, 1968 IN : KWAN, 1977]. Elle a ensuite été développée par FINKELSTEIN ET FAIRLEY, suite aux problèmes d'interprétation des données chiffrées liées aux différents éléments de preuve, dans l'affaire *People v Collins*²⁵ [FINKELSTEIN ET FAIRLEY, 1970].

L'utilisation de ce théorème permet de faire évoluer un rapport de probabilité *a priori* de deux hypothèses compétitives, H_1 et H_2 , vers un rapport de probabilité *a posteriori* de ces deux hypothèses, après l'analyse d'un indice matériel X et d'un échantillon de comparaison provenant d'une source Y. H_1 représente l'hypothèse que la personne mise en cause Y est la source de l'indice matériel X, alors que H_2 représente l'hypothèse que la personne mise en cause Y n'est pas la source de cet indice matériel X ; par définition les hypothèses H_1 et H_2 sont mutuellement exclusives. L'élément de preuve E est le résultat de l'analyse comparative des caractéristiques x de l'indice matériel X avec les caractéristiques y de l'échantillon de comparaison de la source Y. La vraisemblance de E est estimée, d'une part lorsque l'hypothèse H_1 est vérifiée et d'autre part lorsque l'hypothèse H_2 est vérifiée. Le rapport entre ces deux vraisemblances, *likelihood ratio* (LR), est le résultat du calcul de la valeur numérique qui permet de faire évoluer le rapport de probabilité *a priori* vers le rapport de probabilité *a posteriori*.

3.5.4.4.2. Formalisation

$P(H_1)$	Représente la probabilité que l'hypothèse « Y est la source de l'indice matériel X » soit vérifiée, avant l'analyse de x et y.
$P(H_2)$	Représente la probabilité que l'hypothèse « Y n'est pas la source de l'indice matériel X » soit vérifiée, avant l'analyse de x et y.
$\frac{P(H_1)}{P(H_2)}$	Représente le rapport de probabilité <i>a priori</i> des deux hypothèses compétitives H_1 et H_2 , avant l'analyse de x et y.
$P(H_1 E)$	Représente la probabilité que l'hypothèse « Y est la source de l'indice matériel X » soit vérifiée, après l'analyse de x et y.

²⁵ [People v Collins, 68 Cal. 2d 319, 438 P. 2d 33, 66 Cal. Rptr. 497 (1968)]

$P(H_2|E)$ Représente la probabilité que l'hypothèse « Y n'est pas la source de l'indice matériel X » soit vérifiée, après l'analyse de x et y.

$\frac{P(H_1|E)}{P(H_2|E)}$ Représente le rapport de probabilité *a posteriori* des deux hypothèses compétitives H_1 et H_2 , après l'analyse de x et y.

$\frac{P(E|H_1)}{P(E|H_2)}$ Représente le rapport de vraisemblance, *likelihood ratio* (LR), mis en évidence entre le rapport de probabilité *a priori* et le rapport de probabilité *a posteriori*.

$\frac{P(H_1)}{P(H_2)}$	→ multiplié par LR →	$\frac{P(H_1 E)}{P(H_2 E)}$
Rapport de probabilité <i>a priori</i>	$LR = \frac{P(E H_1)}{P(E H_2)}$	Rapport de probabilité <i>a posteriori</i>

Tableau III.1. Schématisation du processus d'inférence de l'identité en sciences forensiques par évaluation de rapports de vraisemblance [CHAMPOD ET TARONI, 1994 ; ROBERTSON ET VIGNAUX, 1995].

3.5.4.4.3. Approches discrète et continue pour l'évaluation de rapports de vraisemblance

La formulation de la méthode d'évaluation de rapports de vraisemblance développée par FINKELSTEIN ET FAIRLEY diffère toutefois de celle initiée par KAPLAN. La probabilité *a priori* que la personne mise en cause soit la source de l'indice matériel, $P(H_1)$, définie par KAPLAN, est équivalente à la *prior probability* de FINKELSTEIN ET FAIRLEY. Par contre la *posterior probability* de FINKELSTEIN ET FAIRLEY ne correspond pas à la probabilité *a posteriori* que la personne mise en cause soit la source de l'indice matériel, $P(H_1|E)$, définie par KAPLAN [KAPLAN, 1968 IN : KWAN, 1977 ; FINKELSTEIN ET FAIRLEY, 1970 ; KWAN, 1977]. En effet, la probabilité *a posteriori*, $P(H_1|E)$, définie par KAPLAN, représente la probabilité d'observer les caractéristiques x et y dans le cas où l'hypothèse H_1 est vérifiée, alors que la *posterior probability*, $P(H_1|E)$ de FINKELSTEIN ET FAIRLEY, représente la probabilité d'une non-discrimination des caractéristiques x et y, en anglais *match*, dans le cas où l'hypothèse H_1 est vérifiée [KAPLAN, 1968 IN : KWAN, 1977 ; FINKELSTEIN ET FAIRLEY, 1970 ; KWAN, 1977].

FINKELSTEIN ET FAIRLEY considèrent donc que l'analyse comparative des caractéristiques x et y aboutit à une décision binaire de discrimination ou de non-discrimination, en anglais *match* ou *non match*, accessible aux méthodes quantitatives non paramétriques d'inférence que sont les tests statistiques de signification de type t-test de Student, test de rang et test de χ^2 . Cette approche peut être qualifiée de discrète puisqu'en cas de correspondance, $P(H_1|E) = 1$. L'approche de KAPLAN, par contre, peut être considérée comme continue car en cas de correspondance $P(H_1|E) \leq 1$ [KWAN, 1977].

En conséquence, la formulation de FINKELSTEIN ET FAIRLEY ne se préoccupe pas de la question de l'intravariabilité de la source pour des raisons de simplicité. Cependant cette

simplification implique des limitations propres aux approches discrètes, souvent mises en évidence, comme le calcul des taux d'erreur de type I et de type II, lié à toute décision binaire, ou le « *fall of the cliff effect* », phénomène décrit par Ken Smalldon [FINKELSTEIN ET FAIRLEY, 1970 ; LINDLEY, 1977 ; AITKEN, 1995 ; EVETT ET BUCKLETON, 1996 ; CURRAN ET AL., 2000].

La formulation de KAPLAN ne souffre pas de ces limitations car, dans l'approche continue, l'intravariabilité de la source Y et l'intervariabilité de l'indice X sont prises en considération pour l'évaluation du rapport de vraisemblance de l'élément de preuve E. Dans le cas où l'intravariabilité et l'intervariabilité sont équivalentes, les probabilités associées à H_1 et à H_2 sont égales et le rapport de vraisemblance vaut 1.

3.5.4.4.4. Conséquences de l'adoption de l'approche par évaluation de rapports de vraisemblance

KAPLAN, ainsi que FINKELSTEIN et FAIRLEY s'accordent pour proposer que la cour soit libre de fixer la valeur de la probabilité *a priori* [KAPLAN, 1968 IN : KWAN, 1977 ; FINKELSTEIN ET FAIRLEY, 1970]. En 1977, LINDLEY formalise ce point de vue : La preuve devrait être présentée au jury sous la forme de déposition au sujet des éléments matériels, X et Y, ainsi que du rapport de vraisemblance qui peut en être déduit. La cour s'appuie sur ce rapport de vraisemblance pour diminuer son incertitude et déterminer le rapport de probabilité *a posteriori* et, delà, la décision d'innocence ou de culpabilité. L'élément de preuve est le travail du témoin, de l'expert ou du scientifique, la décision finale celui du tribunal. Cet auteur soutient que la tâche du scientifique est de déterminer ce que l'élément de preuve signifie en cas de culpabilité et ce qu'il signifie en cas d'innocence [LINDLEY, 1977]. LINDLEY observe aussi l'indépendance de tout ce qui concerne le rapport de probabilité *a priori* vis-à-vis du théorème de Bayes, avant que les éléments de preuve ne soient introduits. Cette constatation est non seulement valable en sciences forensiques, mais aussi en général [LINDLEY, 1977].

3.5.5. Choix d'un processus d'inférence de l'identité

La démonstration de la conformité logique et légale de l'évaluation des rapports de vraisemblance comme méthode quantitative de l'inférence de l'identité en sciences forensiques, effectuée par KWAN en 1977 sur la base des travaux de KAPLAN ainsi que de FINKELSTEIN ET FAIRLEY, a été confirmée à de nombreuses reprises [KAPLAN, 1968 IN : KWAN, 1977 ; FINKELSTEIN ET FAIRLEY, 1970 ; KWAN, 1977 ; KAYE, 1979 ; EVETT, 1987 ; ROBERTSON ET VIGNAUX, 1995]. Pourtant, seuls LEWIS, BROEDERS ainsi que CHAMPOD, DRYGAJLO et MEUWLY, sous l'influence de EVETT, ont suggéré son application à la reconnaissance de locuteurs en sciences forensiques [LEWIS, 1984 ; BROEDERS, 1995 ; MEUWLY ET AL., 1998 ; CHAMPOD ET MEUWLY, 1998 ; MEUWLY, 2000].

La théorie mathématique des probabilités étant cohérente, le théorème de Bayes s'applique tout aussi bien aux probabilités statistiques que subjectives [SAVAGE, 1972 ; DE FINETTI, 1975]. L'évaluation des rapports de vraisemblance peut donc être généralisée à tous les indices, comme le témoignage, l'indice matériel, ou l'aveu, en ce sens qu'elle décrit de façon précise la manière dont ils se combinent, si les juges ou le jury agissent de manière rationnelle.

3.5.5.1. Application à la reconnaissance de locuteurs

3.5.5.1.1. Choix d'une approche

Par sa prise en compte de l'intravariabilité de la source dans l'évaluation de rapports de vraisemblance, l'approche continue initiée par KAPLAN est un processus d'inférence de l'identité plus adapté à la reconnaissance de locuteurs en sciences forensiques que l'approche discrète développée par FINKELSTEIN et FAIRLEY [KAPLAN, 1968 IN : KWAN, 1977 ; FINKELSTEIN ET FAIRLEY, 1970]. En effet, l'évaluation de la variabilité intralocuteur, sur la base de l'analyse de caractéristiques dépendantes du locuteur, représente un enjeu majeur en l'absence de connaissance de caractéristiques spécifiques au locuteur.

3.5.5.1.2. Obtention de l'élément de preuve avec une méthode de reconnaissance automatique de locuteurs

En général, une méthode automatique fournit le résultat de la comparaison d'un modèle de la voix d'un locuteur et d'un échantillon de parole de test sous forme d'un nombre réel, qui représente une distance mathématique ou une proximité statistique, calculée entre le modèle et l'échantillon de test. L'élément de preuve est obtenu de cette manière : il s'agit d'un nombre qui décrit une distance mathématique ou une proximité statistique résultant de la comparaison entre les caractéristiques y du modèle de la voix de la personne mise en cause Y et des caractéristiques x de la voix inconnue enregistrée sur l'indice matériel X .

3.5.5.1.3. Estimation de l'intravariabilité et de l'intervariabilité avec une méthode de reconnaissance automatique de locuteurs

En pratique, ni les enregistrements vocaux ni la méthode de reconnaissance automatique servant à définir l'intravariabilité et l'intervariabilité ne permettent d'obtenir les vraies fonctions de densité de probabilité de la variabilité intralocuteur et interlocuteur, puisque les données ne sont jamais exhaustives et la méthode jamais parfaite. Dès lors, l'approche empirique ne permet par définition qu'une estimation de l'intravariabilité et de l'intervariabilité.

L'estimation de l'intravariabilité de la source Y est obtenue par la comparaison d'un ensemble de modèles de la voix de la personne mise en cause Y avec un ensemble d'échantillons de parole de la personne Y , enregistrés dans différentes conditions. Les distances résultant de ces comparaisons permettent ensuite d'estimer la fonction de densité de probabilité de la variabilité intralocuteur.

L'estimation de l'intervariabilité de l'indice matériel X est obtenue de manière analogue par la comparaison de l'échantillon de parole inconnue avec les modèles des voix de l'ensemble des personnes qui modélisent la population potentielle des auteurs de l'indice matériel X .

3.5.5.1.4. Estimation du rapport de vraisemblance

L'approche empirique n'aboutit qu'à des estimations de l'intravariabilité et de l'intervariabilité ; le rapport de vraisemblance, qui est calculé sur la base de ces estimations, ne peut donc être lui-même qu'une estimation (\hat{LR}). Ce sont la validité des données enregistrées pour l'évaluation de l'intravariabilité et de l'intervariabilité et la fiabilité de la méthode de

reconnaissance utilisée qui déterminent l'adéquation ou la non-adéquation entre le rapport de vraisemblance estimé (\hat{LR}) et le vrai rapport de vraisemblance.

3.5.5.1.5. Formalisation

$P(H_1)$	Représente la probabilité que l'hypothèse « la personne mise en cause Y est effectivement auteur de l'enregistrement présenté comme indice X » soit vérifiée, avant l'analyse de x et y.
$P(H_2)$	Représente la probabilité que l'hypothèse « la personne mise en cause n'est pas auteur de l'enregistrement présenté comme indice X » soit vérifiée, avant l'analyse de x et y.
$\frac{P(H_1)}{P(H_2)}$	Représente le rapport de probabilité <i>a priori posteriori</i> (ou chances <i>a priori</i>) des deux hypothèses compétitives H_1 et H_2 , avant l'analyse de x et y.
$\hat{P}(H_1 E)$	Représente l'estimation de la densité de probabilité de l'élément de preuve E, lorsque l'hypothèse que la personne mise en cause Y est la source de l'enregistrement présenté comme indice X (H_1), est vérifiée.
$\hat{P}(H_2 E)$	Représente l'estimation de la densité de probabilité de l'élément de preuve E, lorsque l'hypothèse que la personne mise en cause Y n'est pas la source de l'enregistrement présenté comme indice X (H_2), est vérifiée.
$\frac{\hat{P}(H_1 E)}{\hat{P}(H_2 E)}$	Représente l'estimation du rapport de probabilité <i>a posteriori</i> (ou chances <i>a posteriori</i>) des deux hypothèses compétitives H_1 et H_2 , après l'analyse de x et y.
$\frac{\hat{P}(E H_1)}{\hat{P}(E H_2)}$	Représente l'estimation du rapport de vraisemblance, <i>likelihood ratio</i> (\hat{LR}), mis en évidence entre le rapport de probabilité <i>a priori</i> et le rapport de probabilité <i>a posteriori</i> .

Prérogative de la cour	Prérogative du scientifique	Prérogative de la cour
$\frac{P(H_1)}{P(H_2)}$	multiplié par \hat{LR} →	$\frac{\hat{P}(H_1 E)}{\hat{P}(H_2 E)}$
Rapport de probabilité <i>a priori</i>	$\hat{LR} = \frac{\hat{P}(E H_1)}{\hat{P}(E H_2)}$	Estimation du rapport de probabilité <i>a posteriori</i>

Tableau III.2. Schématisation du processus d'inférence de l'identité adopté pour la reconnaissance de locuteurs en sciences forensiques

3.6. Évaluation d'une méthode de reconnaissance automatique de locuteurs

La difficulté à définir l'information analysée, l'information dépendante du locuteur, rend la phase d'évaluation d'une méthode de reconnaissance automatique de locuteurs difficile et plus onéreuse que sa phase de mise au point, en termes de moyens et de travail. Trois approches ont été proposées pour l'évaluation de l'efficacité de ces méthodes : l'établissement de modèles théoriques, la comparaison de modèles théoriques et l'évaluation empirique [CAPPE, 1995].

3.6.1. Établissement de modèles théoriques

L'établissement de modèles théoriques peut permettre de dégager de grandes tendances, comme la démonstration que la classification est une tâche plus difficile que la discrimination, dans le cas d'un grand nombre de locuteurs [DODDINGTON, 1985]. La modélisation complète du fonctionnement d'une méthode reste toutefois d'une utilité assez limitée, car en général elle ne correspond que de très loin au fonctionnement en situation réelle [CAPPE, 1995].

3.6.2. Comparaison de modèles théoriques

Il est parfois possible de comparer des techniques sur la base d'arguments théoriques sans recourir à l'expérimentation. Ce type de démarche a notamment été utilisé pour sélectionner les caractéristiques du signal de parole les plus appropriées pour la reconnaissance [DAS ET MOHN, 1971 ; SAMBUR, 1975 ; ATAL, 1976 ; CHEUNG ET EISENSTEIN, 1978]. Cependant, comme le relève avec pertinence CAPPE, il est malheureusement impossible de progresser dans ce domaine sans recours à la modélisation ou à la définition d'hypothèses de travail ne correspondant qu'imparfaitement à la réalité [CAPPE, 1995].

3.6.3. Évaluation empirique

Avec l'arrivée de nouvelles méthodes de reconnaissance au début des années 1980, l'évaluation empirique a supplanté la comparaison des modèles théoriques. Cette évolution s'explique, d'une part, par la difficulté d'analyse théorique de ces méthodes complexes et, d'autre part, par le développement de la micro-informatique, qui a rendu cette évaluation possible [CAPPE, 1995]. L'évaluation empirique constitue une méthode de validation très satisfaisante car elle permet d'obtenir directement une estimation de la fiabilité en situation réelle. Dans le cas forensique cette phase d'évaluation consiste à observer l'adéquation entre l'estimateur du rapport de vraisemblance (\hat{LR}) et la réalité. Cette stratégie est en ce sens beaucoup plus efficace que les arguments théoriques, qui ne peuvent être utilisés que pour comparer différentes méthodes entre elles. Toutefois, ce caractère empirique limite l'interprétation et le domaine de validité des résultats aux enregistrements de qualité comparable à celle des enregistrements utilisés dans la phase d'évaluation.

3.6.4. Choix d'une méthode d'évaluation

L'évaluation empirique d'une méthode de reconnaissance automatique de locuteurs en vue de son application forensique se révèle particulièrement difficile, puisque la maîtrise des paramètres qui conditionnent la qualité des enregistrements présentés comme indices est inexistante, à l'exception de ceux concernant le système d'enregistrement²⁶. Elle constitue néanmoins le meilleur moyen d'estimer les performances du système développé dans le cadre de cette recherche.

3.6.4.1. Critères de sélection des bases de données

Une procédure de reconnaissance automatique de locuteurs nécessite la constitution de deux bases de données. La première sert à estimer la variabilité interlocuteur à l'intérieur de la population des locuteurs qui sont potentiellement à l'origine de l'enregistrement considéré comme indice. La seconde, de plus petite taille, permet l'estimation de la variabilité intralocuteur de la ou des personne(s) suspectée(s) d'être la source de l'indice. En sciences forensiques, ces concepts d'intravariabilité et d'intervariabilité ont été initiés par TIPPET ET AL., dans le domaine de l'interprétation des résultats d'analyse de peintures automobiles, sous les dénominations respectives de *within source comparison* et de *between source comparison* [TIPPET ET AL., 1968].

3.6.4.1.1. Détermination de la langue parlée

L'écoute de l'enregistrement considéré comme indice permet de déterminer la langue parlée et l'accent régional du locuteur inconnu, de manière subjective mais fiable²⁷; dans une moindre mesure le sexe de la personne inconnue peut être déterminé, notamment sur la base de la hauteur de la fréquence fondamentale de sa voix²⁸. D'autres critères de qualification, fondés sur la qualité de la voix ou sur une proximité auditive, peuvent être envisagés à petite, mais pas à grande échelle; de plus ces critères restent difficiles à systématiser.

3.6.4.1.2. Estimation de la variabilité interlocuteur

Les critères, mis en évidence lors de l'écoute initiale de l'indice, servent à définir la population des locuteurs qui en sont potentiellement l'origine et à sélectionner une fraction de ces personnes pour modéliser cette population d'auteurs potentiels. La qualité de la modélisation dépend de la taille de la base de données et de la justesse avec laquelle celle-ci représente la population potentielle.

Le rôle de cette première base de données consiste à mesurer la variabilité interlocuteur, c'est-à-dire à calculer la distance mathématique ou la proximité statistique entre l'indice matériel et la voix des locuteurs de la population potentielle. Cette estimation empirique est réalisée en

²⁶ *supra* : 2.3. Collecte de l'indice matériel

²⁷ *supra* : 2.3.7.1. Influence de l'investigation préliminaire

²⁸ *infra* : 4.3.3.3.6. Détermination du genre du locuteur

comparant la voix de l'indice avec chacun des modèles des voix des locuteurs de la première base de données.

3.6.4.1.3. Estimation de la variabilité intralocuteur

Cette base de données est constituée des enregistrements de la personne suspectée d'être la source de l'indice. Le rôle de cette base de données consiste à mesurer la variabilité intralocuteur de la personne mise en cause, c'est-à-dire à mesurer la distance mathématique ou la proximité statistique des énoncés de cette personne avec les modèles de sa propre voix. Deux types d'enregistrement lui sont demandés : Premièrement l'enregistrement de plusieurs sessions en tous points analogues aux sessions existant dans la base de données interlocuteur, si possible réparties sur une période de temps comparable à celle de l'affaire, pour permettre une évaluation de la variabilité intralocuteur sur la même durée ; deuxièmement, l'enregistrement d'une longue session, de 5 à 15 minutes selon les locuteurs, de manière à modéliser la variabilité intralocuteur dans différentes situations et divers styles d'élocution.

3.6.4.1.4. Constitution d'enregistrements de test

Pour les besoins de cette recherche, les personnes sélectionnées pour jouer le rôle des personnes mises en cause ont aussi contribué à constituer un ensemble d'enregistrements de test, simulant les indices qui peuvent être rencontrés en cas d'abus de téléphone ou de mesure de surveillance.

3.7. Conclusion

Premièrement, cette analyse méthodologique met en évidence la nécessité de présenter l'état de l'art dans le domaine de la reconnaissance de locuteurs en sciences forensiques. Elle montre deuxièmement que la démarche scientifique est une démarche appropriée pour parvenir à cette présentation de l'état de l'art. Troisièmement, l'analyse des différentes méthodes pratiquées pour la reconnaissance de locuteurs en sciences forensiques conduit à considérer l'approche automatique d'un point de vue théorique et expérimental, plutôt que les approches auditive ou spectrographique. Quatrièmement, le résultat de l'étude des différents processus d'inférence de l'identité envisagés pour la reconnaissance de locuteurs en sciences forensiques indique la conformité logique et légale de l'approche par évaluation de rapports de vraisemblance. Finalement, l'évaluation empirique est considérée comme le meilleur moyen d'estimer les performances du système de reconnaissance automatique de locuteurs développé dans le cadre de la recherche théorique et expérimentale.

PARTIE 2

RECHERCHE BIBLIOGRAPHIQUE

IV. APPROCHE AUDITIVE

4.1. La perception de la voix et de la parole

4.1.1. Principes de la perception

L'anatomie et la physiologie de l'oreille humaine sont en premier lieu adaptées à la perception de la voix humaine. La perception de la parole est généralement décrite comme un processus comprenant plusieurs étages d'analyse dans la transformation de la parole en message [STUDDERT-KENNEDY, 1974 ; STUDDERT-KENNEDY, 1976]. Bien que la nature exacte de chacun des étages décrits et leurs interactions ne soient encore que supposés, ils sont justifiables d'un point de vue linguistique. Sur la base des études de STUDDERT-KENNEDY, PISONI ET LUCE proposent cinq étages conceptuels d'analyse : l'analyse auditive périphérique, l'analyse auditive centrale, l'analyse acoustique phonétique, l'analyse phonologique et les analyses d'ordre plus élevé : lexicale, syntaxique et sémantique [STUDDERT-KENNEDY, 1974 ; PISONI ET LUCE, 1987].

4.1.2. Le processus de discrimination et d'identification de locuteurs

Les processus de discrimination et d'identification de locuteurs par l'être humain sont souvent présentés comme des variantes d'un processus cognitif unique [HECKER, 1971 ; BRICKER ET PRUZANSKY, 1976]. Cependant, l'étude des performances de patients atteints de troubles de discrimination ou d'identification de locuteurs tend à montrer que ces deux tâches procèdent de fonctions neuropsychologiques différentes, sous-tendues par des régions cérébrales distinctes. Les performances aux tests de discrimination et d'identification ne sont que modérément corrélées chez les sujets normaux [ROSE ET DUNCAN, 1995] et ne sont pas corrélées significativement chez des patients porteurs de lésions cérébrales unilatérales droite ou gauche. VAN LANCKER montre que les patients cérébrolésés unilatéralement à droite ou à gauche présentent des troubles de la discrimination des voix, alors que seuls les patients atteints de lésions unilatérales droites présentent des difficultés lors du test d'identification de locuteurs familiers, par rapport au groupe témoin [VAN LANCKER ET AL., 1987]. Dans cette même étude, 44 % des patients cérébrolésés obtiennent des résultats significativement différents pour les deux tâches. De plus, la discrimination ne constitue pas une première étape vers l'identification, puisque certains patients sont capables d'identifier les locuteurs tout en échouant au test de discrimination ; le phénomène opposé est observé chez d'autres. Quatre patients présentant des déficits importants, soit de l'identification, soit de la discrimination, ont également été testés dans les épreuves de discrimination et d'identification de visages, de bruits et de sons de l'environnement. Les perturbations de la perception des voix semblent spécifiques, car elles ne sont pas corrélées à d'autres déficits dans les différents tests proposés.

Une analyse tomographique a permis de mettre en évidence que les patients présentant un déficit de l'identification des voix familières sont atteints de lésions pariétales droites, alors que les

patients atteints de troubles de la discrimination de voix non familières sont atteints de lésions touchant le lobe temporal droit ou gauche [VAN LANCKER *ET AL.*, 1988].

Ces résultats peuvent être interprétés comme une différence de stratégie entre la tâche de discrimination et la tâche d'identification de locuteurs. La tâche de discrimination semble être principalement l'œuvre d'une stratégie méthodologique de comparaison, fondée sur l'analyse de caractéristiques et de paramètres acoustiques de base, centrés dans la mémoire à court terme, comparable à l'analyse acoustique phonétique, catégorielle [VAN LANCKER *ET AL.*, 1987 ; VAN LANCKER *ET AL.*, 1989].

La tâche d'identification, par contre, procéderait d'une stratégie cognitive de reconnaissance de formes, basée sur l'appariement de structures holistiques, résidant dans la mémoire à long terme, à rapprocher de l'analyse sémantique. Un modèle actuel de la spécialisation des hémisphères cérébraux associe d'ailleurs les tâches de reconnaissance de formes à des régions cérébrales centrées dans l'hémisphère droit et les processus analytiques à des régions cérébrales centrées dans l'hémisphère gauche [BRYDEN, 1982 ; BRADSHAW ET NETTLETON, 1983].

Cette interprétation de la dissociation entre discrimination et identification n'est cependant ni définitive ni catégorique, car la réalisation de ces deux tâches suppose, d'une part, un processus de traitement en deux grands niveaux, reposant sur de multiples séquences et opérations interactives, et, d'autre part, la nécessité d'un traitement massivement parallèle de l'information auditive [WATROUS, 1990 ; EUSTACHE, 1995].

4.2. Les méthodes de reconnaissance auditive

L'étude de la reconnaissance de locuteurs par audition se concentre sur l'étude de la manière dont les auditeurs humains réalisent la tâche d'association d'une voix particulière à un individu particulier ou à un groupe et notamment dans quelles circonstances une telle tâche peut être remplie [NOLAN, 1983].

En sciences forensiques, la reconnaissance de locuteurs par l'audition est pratiquée soit par des experts, phonéticiens ou spécialistes des sciences de la parole, sur la base de principes scientifiques, soit de manière perceptive par des profanes, principalement les victimes ou les témoins d'une infraction [KÜNZEL, 1994B].

4.3. Procédure de reconnaissance par des profanes

Dans le cadre du procès pénal, le « principe d'individualité »²⁸ de la voix humaine a été implicitement accepté dès le XVII^e siècle. Le premier cas d'identification de la voix d'un suspect par un témoin remonte à 1660, lors du procès de l'un des vingt-neuf hommes jugés pour haute trahison

²⁸ *supra* : 1.2.2. La voix comme moyen d'identification

ayant conduit à l'exécution de Charles I^{er} d'Angleterre ²⁹ [BOLT ET AL., 1979]. Cependant, ce n'est qu'à la fin du XIXe siècle que la valeur de la voix, en tant que caractère d'identité, a été étudiée.

4.3.1. Approche descriptive

4.3.1.1. L'anthropométrie ou bertillonnage

Dès 1879, Alphonse BERTILLON propose une méthode de reconnaissance des récidivistes basée sur les mensurations de certaines longueurs somatiques particulièrement invariables : le signalement anthropométrique [BERTILLON, 1881]. Il s'assortit d'une série de systèmes complémentaires, dont l'ensemble a reçu du professeur Alexandre Lacassagne le nom de bertillonnage. Parmi les nombreux caractères d'identité proposés à cette époque en police scientifique, certaines particularités d'ordre physiologique, telles que la démarche, l'allure, le geste, le regard et la parole sont inclus dans le signalement anthropométrique [LOCARD, 1909].

La voix et le langage sont caractérisés pour la première fois dans les instructions signalétiques de BERTILLON :

«Le timbre de la voix est l'un des caractères les plus distinctifs de l'individualité. Chacun sait que nous reconnaissons nos parents, nos amis, toutes les personnes avec lesquelles nous sommes en rapport journalier, à distance, d'une pièce à une autre, rien qu'au son de leur voix. Malheureusement, le phonographe mis à part, aucun signe n'est plus difficile à noter. On signalera les voix particulièrement graves ou aiguës, la voix de fausset, la voix féminine chez l'homme et la voix masculine chez la femme. Les principaux vices organiques d'articulation sont : le zéaiement, le chuintement, le bégaiement et le grasseyement.

La connaissance raisonnée des différents accents qui caractérisent chacune des provinces de la France serait certes d'une grande utilité pour l'identification des inconnus qui cachent leur nom, si en cette matière si délicate la théorie pouvait suppléer à la pratique. La distinction des principaux accents étrangers, pour peu qu'on ait eu l'occasion d'y familiariser son oreille, est certes plus aisée et plus tranchée que celle des accents provinciaux. Chaque nationalité transporte dans sa manière de parler une langue étrangère, la prononciation, les règles de grammaire et les tournures de phrase usitées en sa propre langue » [BERTILLON, 1893].

BERTILLON postule que la voix répond au « principe d'individualité », mais il souligne la faible capacité de l'être humain à décrire une voix, ainsi que le manque de connaissances théoriques en linguistique et en phonétique. Il tente aussi une première analyse en distinguant dans le signal de parole les éléments de phonétique acoustique comme le timbre, les éléments de phonétique fonctionnelle, telle la prononciation et ses défauts, et les éléments de linguistique, comme la langue et ses particularités.

Le dénuement des caractérisations de la voix, présentées dans la télégraphie chiffrée du portrait bertillonnien d'Archibald Rodolphe REISS, « Le Portrait Parlé », illustrent d'ailleurs ce manque de connaissances théoriques et cette difficulté à décrire la voix (Tableau IV.1.) [REISS, 1907] :

²⁹ [Hulet's trial, 5 Howell's St. Trials 1185, 1187 (1660)]

0,87 : La voix	0,871 : voix grave	0,875 : zézaiement
	0,872 : voix aiguë	0,876 : chuintement
	0,873 : voix féminine	0,877 : bégaiement
	0,874 : voix masculine	0,878 : accent étranger

Tableau IV.1. Les caractérisations de la voix dans « Le Portrait Parlé » de REISS

Cette inaptitude des locuteurs à augmenter, par eux-mêmes, le nombre d'adjectifs permettant de caractériser les voix a d'ailleurs été mise en évidence expérimentalement [VOIERS, 1964].

4.3.1.2. Le signalement descriptif fonctionnel

Dans son « *Trattato di Polizia Scientifica* », Salvatore OTTOLENGHI souligne qu'un individu ne se reconnaît pas seulement par ses attributs anatomiques, mais aussi par la façon qu'il a de se présenter, par sa démarche, par sa voix, par sa mimique, par son écriture, par sa force, par son acuité sensorielle et par ses attitudes organiques viscérales [OTTOLENGHI, 1910]. Avec beaucoup de détails, il analyse cette partie du signalement qu'il appelle « signalement descriptif fonctionnel » et montre qu'en plus de l'observation de l'anatomie, celle de la physiologie peut contribuer à l'identification d'une personne :

«Le signalement physique doit tenir compte de la propriété que l'individu a de parler, car le langage peut offrir des caractéristiques de sa personnalité très importantes et faciles à relever. Pour comprendre les caractères du langage, il faut en connaître les mécanismes de formation. On ne doit pas s'occuper ici du contenu du langage, mais de la formation des mots et de leur émission.

Nous parlons surtout, car nous entendons. C'est-à-dire que des sons parviennent à notre ouïe. Ils sont récoltés alors par notre organe auditif, puis transmis au centre cortical (première circonvolution temporale) où s'accumulent ces images acoustiques. Quand on parle, on répète en fait des sons que l'on a entendus une fois (le sourd-muet ne parle pas, car il n'entend pas). Et qui a parlé et perd la faculté d'entendre, perd nécessairement la faculté de parler.

Il n'est pas suffisant d'entendre, il faut former le mot et le prononcer. La formation de la parole est un mécanisme qui se met en place grâce à la participation du centre psychomoteur de Broca, qui se trouve à la base de la troisième circonvolution frontale et des zones psychomotrices adjacentes. Il s'agit de l'évocation des images verbales sensorielles accumulées dans les centres psychosensoriels, en particulier dans le centre auditif, et dans la formation des images motrices relatives. La parole formée dans le centre cortical est transformée en acte moteur à travers les systèmes centraux et périphériques qui sont responsables de la coordination des mouvements phonatoires. Cette transmission se fait au moyen du système nerveux moteur, localisé entre les centres corticaux et les muscles destinés à l'articulation de la parole. Il s'ensuit la formation et l'émission des sons dans les systèmes phonatoires externes : larynx, pharynx, bouche et nez.

Pour étudier les fonctions essentiellement motrices, il faut s'occuper seulement du mécanisme intrinsèque de la parole, c'est-à-dire de la manière avec laquelle l'individu émet des sons vocaux (phonation), de celle dont il prononce les mots et de celle dont il les articule (articulation). On s'occupera des autres fonctions du langage dans les examens psychologiques » [OTTOLENGHI, 1910].

L'auteur mentionne que la production de la parole est le résultat des deux fonctions mécaniques de base que sont la phonation et l'articulation et, comme BERTILLON, fait la distinction entre analyse de la voix et analyse du langage. Sa description de la voix est beaucoup plus fouillée que celle de BERTILLON, car elle fait référence à de nombreuses connaissances médicales, qui rappellent l'origine professionnelle d'OTTOLENGHI :

«De la voix, il faut considérer le volume ou la force, la hauteur du son ou le ton, la qualité du son ou le timbre, l'agilité, le type et l'intonation. Selon la force, la voix peut être forte, moyenne ou faible. Elle peut, dans des cas exceptionnellement morbides, être très forte, très faible ou manquer (pseudo-mutisme ou aphonie). L'aphonie peut être temporaire ; c'est le cas des hystériques, qui peuvent tout à coup perdre la voix.

Selon la hauteur ou le ton, la voix peut être plus ou moins haute ou plus ou moins basse. Selon le son, on distingue la voix ordinaire, la voix nasale et la voix gutturale (un peu rauque). Certaines voix ont le caractère de véritables marques personnelles, car elles contiennent des sons spéciaux qui dépendent de différentes causes, par exemple la voix stridente des tuberculeux, qui ont des processus maladifs à la muqueuse du larynx, la voix toute spéciale, presque aphone, de ceux qui ont le bec de lièvre, la gorge de loup, la paralysie du vélus palatal pendant, des maladies ou des déformations de la cavité nasopharyngale.

Selon le timbre, la voix peut être plus ou moins claire, rugueuse ou stridente ; selon la malléabilité ou l'agilité, elle peut être agile, fluide ou tremblante. Le type ou la qualité de la voix peut varier selon l'âge ou le sexe. On parlera de voix infantile, sénile, de voix masculine ou féminine. Les caractères distinguant ces divers types de voix sont trop connus pour être décrits. Le type de voix spéciale de l'eunuque complète le type eunuchoïde déjà décrit. Des variations très exceptionnelles se retrouvent seulement dans certains types d'aliénation mentale dans lesquels la voix perd son type humain : émission de cris plus ou moins aigus, de hurlements comme des vaches mugissantes, de grognements, de miaulements, de bêlements ou d'abolements.

L'intonation de la voix varie selon le renforcement ou l'abaissement de la voix, la modulation selon l'accentuation des syllabes, les pauses, le ton différent, la cadence et la rapidité. La voix peut souvent changer de force, elle peut passer d'un excès à l'autre. La phrase peut être plus ou moins modulée, afin d'être harmonieuse ou rugueuse, uniforme ou non. L'accentuation des syllabes peut être normale, absente, exagérée, ce qui équivaut à souligner certains mots dans l'écriture. Les pauses entre un mot et l'autre peuvent être uniformes ou non, proportionnées ou pas. Le ton peut être varié, il peut être exalté, déprimé ou ordinaire. La cadence peut être plus ou moins soutenue, musicale ou monotone. Selon la rapidité de la phonation les mots se suivent, rapides ou lents » [OTTOLENGHI, 1910].

OTTOLENGHI met déjà en évidence certaines causes de la variabilité du signal de parole, qu'elle soient dues au locuteur, comme l'âge ou le sexe, ou au message lui-même, comme le ton ou l'accentuation. Il met aussi en évidence la valeur informative des silences et établit une première comparaison entre le signal de parole et l'écriture.

Dans la suite de sa description, il énumère toutes sortes de défauts de prononciation et d'articulation et les maladies ou malformations qui leur sont liées. La plupart ne sont cependant plus d'actualité. Les observations d'OTTOLENGHI sur les handicaps cérébro-moteurs illustrent par

contre la difficulté à appréhender les mécanismes de production de la parole qui, aujourd'hui, restent encore en grande partie à découvrir :

«On ne peut pas se désintéresser des altérations de la formation de la parole, qui dépendent de lésions des centres mnémoniques du cortex et qui peuvent fournir des caractéristiques personnelles très utiles. Ces lésions s'appellent dysphasies et consistent dans la perte de la parole ou dans le fait que les images verbales deviennent floues. Parmi ces dysphasies, l'aphasie motrice est très intéressante : l'individu veut commencer à parler spontanément ou répéter des mots dits par autrui, mais il ne peut que partiellement ou pas du tout s'exprimer. Il intervertit une syllabe ou un mot, car le mécanisme d'évocation des images motrices de la parole est défectueux. Parfois, ce défaut est limité à quelques mots, par exemple aux noms. Il se peut que la formation de la parole soit très simple ; par conséquent elle est émise impulsivement avec précipitation et on aura le contraire de l'aphasie, c'est-à-dire la loquacité exagérée qui s'appelle soliloque ou logorrhée. On peut avoir la répétition insistante de certains mots, ce qui s'appelle écholalie, ou de certains noms (onomatopée) ou de nombres (arithmomanie) » [OTTOLENGHI, 1910].

4.3.2. Limites de l'approche descriptive

Dans « Les Preuves de l'Identité », Edmond LOCARD reprend les caractérisations établies par OTTOLENGHI, mais souligne que toute description d'une perception auditive est entachée par la subjectivité :

«La méthode est terriblement incertaine. Il faut tenir compte des conditions dans lesquelles le témoin écoute. Or ces conditions ne sont pas précisément excellentes. D'une part, les voix entendues risquent de ne pas être à leur diapason normal. Un homme qui menace ou qui frappe, un autre qu'on égorge, ne parle pas sur le ton de commérage quotidien, ni sur celui d'une discussion d'une société savante. D'autre part, celui qui écoute est vraisemblablement fort troublé, soit que la scène dont il est témoin auriculaire le remplisse d'horreur, soit qu'il craigne pour sa propre vie. Or l'émotion a pour résultat de troubler la perception et de rendre les souvenirs non seulement imprécis, mais informes ou inexacts » [LOCARD, 1932].

LOCARD met en évidence l'état psychologique particulier de l'auteur, au moment de l'acte délictueux, comme cause de variabilité du signal sonore et celui du témoin, comme cause de reconnaissance incorrecte. Dans « L'Enquête Criminelle », il analyse l'influence de la subjectivité sur la description des perceptions auditives et mentionne l'importance de la psychologie expérimentale pour évaluer les différentes perceptions humaines :

«Avec les sensations auditives nous entrons dans la partie utile du témoignage. Encore faut-il distinguer la perception des sons ou des bruits de celle de la voix parlée. Dans la multitude des ondes qui, transmises par le tympan et la chaîne des osselets, éveillent des vibrations dans les arcs de Corti, un petit nombre parviennent jusqu'au champ de la conscience : celles que l'attention choisit parce qu'elles sont utiles et celles que l'attention subit parce qu'elles sont anormales. Mais il s'en faut que les qualités de ces ondes soient directement perçues. Entre la sensation brute et l'image qui tendra à recevoir une fixation, d'inévitables différences se produisent. Le phénomène acoustique ne s'enregistre qu'après une assimilation déformante constituée par le raisonnement, aussi inconscient et sommaire que l'on voudra, mais dont l'existence est constamment démontrable » ... «Distinguer une voix est une opération physique

complexe qui comporte la perception des trois éléments : hauteur, intensité et timbre, leur comparaison avec des éléments identiques déjà perçus, et l'affirmation de cette identité. On sent combien une telle opération comporte de difficultés, si les conditions ne sont pas absolument favorables, s'il s'agit par exemple de voix chuchotée ou de voix déguisée. Néanmoins lorsque la personne qu'il s'agit d'identifier est de l'intimité du témoin, et que celui-ci n'est pas stupide, la réponse peut être intéressante » [LOCARD, 1932].

4.3.3. Approche expérimentale

4.3.3.1. Méthodologie

La performance humaine dans la tâche de reconnaissance auditive de locuteurs a été évaluée lors de nombreuses expériences de psychologie expérimentale. Comme la plupart des études menées sont originales du point de vue méthodologique, les résultats obtenus sont souvent difficilement comparables. Ils permettent néanmoins d'évaluer la validité des résultats obtenus lors de procédures d'écoute et de reconnaissance par les victimes ou les témoins d'une infraction en sciences forensiques.

BRICKER ET PRUZANSKY ont proposé une grille d'analyse des différentes variables de la procédure expérimentale, en calquant leur modèle sur les différents composants du processus de reconnaissance auditive de locuteurs. Chacune de ces variables est déterminée par l'expérimentateur selon le but de son expérience (Tableau IV.2.) [BRICKER ET PRUZANSKY, 1976].

Composants du processus de reconnaissance	Locuteur	Énoncés	Transmission orale - auditive	Processus sensoriels et perceptifs	Processus de décision
Variables de la procédure expérimentale	Ensemble de référence des locuteurs	Caractéristiques du message	Caractéristiques du canal de transmission	Auditeurs	Tâche et mesure de performance

Tableau IV.2. Grille d'analyse des différentes variables de la procédure expérimentale

4.3.3.2. Tâche et mesure de performance

Selon les définitions proposées par BRADSHAW et NETTLETON ainsi que VAN LANCKER³⁰, la tâche effectuée lors de procédures d'écoute est une tâche de classification lorsqu'il est demandé aux auditeurs de faire appel au souvenir qu'ils ont de la voix entendue au moment de la commission de l'infraction [BRADSHAW ET NETTLETON, 1983 ; VAN LANCKER ET AL., 1987]. Par contre, il s'agit d'une tâche de discrimination lorsqu'il leur est demandé de comparer un enregistrement de parole inconnue à la voix de différents locuteurs.

³⁰ *supra* : 4.1.2. Le processus de discrimination et d'identification de locuteurs

4.3.3.3. Expériences

4.3.3.3.1. Ensemble de référence des locuteurs et auditeurs (Tableau IV.3.)

BRICKER ET PRUZANSKY montrent que, dans un groupe de dix locuteurs familiers des auditeurs, le taux d'identification à partir de phrases est d'environ 98 % [BRICKER ET PRUZANSKY, 1966]. Lorsque ce groupe est constitué de locuteurs ayant des voix similaires, il diminue à 85 % et à 66 % pour des locuteurs non familiers des auditeurs [ROSE ET DUNCAN, 1995]. Lorsque la taille de l'ensemble de référence augmente de six à dix locuteurs non familiers et ayant des voix similaires, le taux d'identification chute encore de 62 % à 40 % [WILLIAMS, 1964]. Les conditions de cette dernière expérience correspondent bien aux conditions forensiques.

La mise en cause de personnes est le résultat du travail des personnes chargées de l'enquête ; le nombre de personnes mises en cause varie selon les circonstances du cas, mais ce tri *a priori* favorise la présence de voix similaires dans le groupe des personnes mises en cause. Aucun degré de familiarité entre auditeur et locuteur ne peut être envisagé *a priori*, car seule une identification de l'auteur par le témoin ou la victime au moment de l'acte délictueux permettrait de l'établir, ce qui rendrait caduque la démarche de reconnaissance de locuteurs [BROEDERS, 1996].

La reconnaissance auditive de locuteurs ne dépend donc pas seulement des caractéristiques individuelles de chacun des locuteurs, mais aussi de celles des locuteurs de l'ensemble de référence et de la taille de cet ensemble [WILLIAMS, 1964]. Une étude comparative de HOLLIEN confirme ces résultats. Les performances des auditeurs familiers des locuteurs sont supérieures à celles des auditeurs non familiers, 98 % contre 40 %, et elles diminuent encore à 27 %, lorsque les auditeurs non familiers ne comprennent pas la langue des locuteurs [HOLLIEN ET AL. 1982]. Comme plusieurs autres travaux, cette étude met en évidence la grande variabilité des performances individuelles des auditeurs [STEVENS ET AL., 1968 ; ROSENBERG, 1973 ; SCHMIDT-NIELSEN ET STERN, 1985]. ATWOOD ET HOLLIEN ont cependant montré qu'en moyenne les auditeurs sous le coup d'une émotion reconnaissent mieux les locuteurs que les autres, contrairement à ce que pensait LOCARD [LOCARD, 1932 ; ATWOOD ET HOLLIEN, 1986]. Il ne semble pas qu'un entraînement spécifique permette d'améliorer les performances des auditeurs, mais CLIFFORD a tout de même mis en évidence que les performances d'auditeurs aveugles sont, en moyenne, de 25 % supérieures à celles d'auditeurs jouissant de la vue [CLIFFORD ET AL. 1981].

Lors de la commission d'une infraction, l'auditeur, victime ou témoin, a la plupart du temps son attention perturbée et ne peut mémoriser qu'incidemment une voix. CLIFFORD et MCCARDLE montrent que la voix d'un locuteur mémorisée incidemment est identifiée parmi dix locuteurs dans 60% des cas après un jour [CLIFFORD ET MCCARDLE, 1980 IN : CLIFFORD, 1980]. Ces résultats vont dans le même sens que l'étude de HINTZMAN et celle de LIGHT, qui montrent qu'à long terme, les performances d'identification d'une voix mémorisée incidemment ne sont pas meilleures que celles résultant d'un choix effectué au hasard [HINTZMAN ET AL. ; 1972 ; LIGHT ET AL. ; 1973]. De même, les auditeurs reconnaissent mieux la voix d'un partenaire de conversation que s'ils écoutent passivement deux interlocuteurs sans prendre part à leur conversation [HAMMERSLEY ET READ, 1985 ; VAN LANCKER ET AL., 1985A ; VAN LANCKER ET AL., 1985B].

Auteurs	Ensemble de référence des locuteurs	Caractéristiques du message	Caractéristiques du canal de transmission	Auditeurs	Tâche et mesure de performance
BRICKER ET PRUZANSKY (1966)	10 locuteurs familiers	Phrase	Haute qualité	16	Identification : 98 %
ROSE ET DUNCAN (1995)	6 locuteurs ayant des voix similaires : A : 4 familiers B : 2 non familiers	10 s de parole spontanée	Haute qualité	10	Identification : A : 85 % B : 66 %
WILLIAMS (1964)	Locuteurs non familiers ayant des voix similaires A et B : 6 locuteurs C : 10 locuteurs	Phrase prononcée normalement et chuchotée	Haute qualité	36	Identification : A : 62 % B : 50 % C : 40 %
HOLLIEN ET AL. (1982)	10 locuteurs	Phrases de 50 à 58 mots, tirées du texte : « My Grandfather »	Haute qualité	A : 10 auditeurs familiers B : 47 auditeurs non familiers C : 14 auditeurs non familiers de langue étrangère	Identification A : 98 % (90 - 100 %) B : 40 % (5 - 80 %) C : 27 % (15 - 45 %)
CLIFFORD ET DENOT (1980) IN : CLIFFORD (1980)	10 locuteurs	Une phrase de 29 syllabes	Haute qualité	15 auditeurs (Mémoire secondaire)	Identification après un jour : 60.4 %

Tableau IV.3. Influence de l'ensemble de référence des locuteurs et auditeurs

Selon CLIFFORD, les performances des auditeurs de vingt à quarante ans sont supérieures à celles d'auditeurs âgés de plus de cinquante ans, lors d'expériences portant sur la taille de l'ensemble des locuteurs, 44 % contre 32 %, sur le déguisement de la voix, 30 % contre 20 %, et lors d'expériences d'identification en ensemble ouvert ou fermé, 59 % contre 49 % [CLIFFORD, 1980].

4.3.3.3.2. Influence d'une modification de la voix (Tableau IV.4. et tableau IV.5.)

Le déguisement, modification volontaire de la voix, influence les performances auditives de reconnaissance de locuteurs, mais la dégradation observée dépend de la stratégie de déguisement choisie ; elle s'étend de 22 % pour l'élocution lente à 33 % pour la voix hypernasale dans l'expérience de REICH ET DUKE [REICH ET DUKE, 1979].

Les modifications involontaires de la voix, comme les conditions de stress, la fatigue, une émotion forte ou le changement de ton dans l'élocution, pénalisent aussi très fortement la performance. Celle-ci diminue de plus de moitié dans plusieurs expériences [SASLOVE ET YARMEY, 1980 ; HOLLIEN ET AL., 1982 ; HOMAYOUNPOUR ET AL., 1993], et de 50 à 33 % dans l'expérience de CLIFFORD ET DENOT [CLIFFORD ET DENOT IN : CLIFFORD, 1980]. Les émotions, par exemple, influencent la hauteur et la variabilité de la fréquence fondamentale, la tessiture, la position des formants, l'intensité et le tempo [WILLIAMS ET AL., 1970 ; LEVIN ET LORD, 1975 ; KRAUSE, 1976 ; SCHERER, 1981]. La taille de l'ensemble des locuteurs et l'âge des auditeurs influencent aussi les résultats, lorsque la voix est modifiée. Le taux d'identification passe de 36 % à 17 % en cas

d'augmentation de quatre à huit locuteurs et les performances du groupe d'auditeurs âgés de plus de cinquante ans sont moins élevées que celles des groupes d'auditeurs de seize à vingt ans et de vingt à cinquante ans qui, elles, sont équivalentes [CLIFFORD, 1980].

Auteurs	Ensemble de référence des locuteurs	Caractéristiques du message	Caractéristiques du canal de transmission	Auditeurs	Tâche et mesure de performance
REICH ET DUKE (1979)	40 locuteurs (21 - 42 ans)	Phrases : A : v. non déguisée B : v. vieillie C : v. rauque D : v. hypernasale E : élocution lente F : déguisement libre	Haute qualité (chambre sourde)	1. : 24 auditeurs non experts 2. : 24 auditeurs experts	Discrimination : A1 : 92 % A2 : 92 % B1 : 68 % B2 : 80 % C1 : 68 % C2 : 81 % D1 : 59 % D2 : 72 % E1 : 70 % E2 : 79 % F1 : 61 % F2 : 74 %
CLIFFORD (1980)	A : 4 locuteurs B : 6 locuteurs C : 8 locuteurs	Une phrase avec voix déguisée comparée à une phrase avec voix normale prononcée par A, B, C	Haute qualité	108 auditeurs et auditrices : 1 : 16 - 20 ans 2 : 20 - 40 ans 3 : 40 - 80 ans	Identification de la voix déguisée : A1 : 42 % B1 : 32 % C1 : 17 % A2 : 42 % B2 : 28 % C2 : 17 % A3 : 25 % B3 : 17 % C3 : 18 %
HOLLIEN ET AL. (1982)	10 locuteurs	Phrases de 50 à 58 mots, tirées du texte : « My Grandfather » A : Voix normale B : Condition de stress C : Déguisement	Haute qualité	1 : 10 auditeurs familiaux 2 : 47 auditeurs non familiaux 3 : 14 auditeurs non familiaux de langue étrangère	Identification : A1 : 98 % A2 : 97.5 % A3 : 79 % B1 : 40 % B2 : 34 % B3 : 21 % C1 : 27 % C2 : 27 % C3 : 18 %
HOMAYOUNPOUR ET AL. (1993)	12 locuteurs de 24 à 55 ans	52 paires de phrases de 10 s A : Voix normale B : Sentiment de bonheur, de colère ou fatigue	Haute qualité	69 auditeurs non familiaux	Discrimination A : 76 % (faux rejet : 11 % ; fausse acceptation : 13 %) B : 34 % (faux rejet : 48 % ; fausse acceptation : 20 %)

Tableau IV.4. Influence de la modification de la voix

La reconnaissance auditive de voix imitées et de voix de jumeaux a été peu abordée. En cas d'imitation, ROSENBERG montre que la performance des auditeurs dépend des qualités de l'imitation et que certains auditeurs se laissent beaucoup plus facilement abuser que d'autres [ROSENBERG, 1973]. L'expérience de HOMAYOUNPOUR et ses collègues, avec des imitateurs non professionnels, confirme ces résultats [HOMAYOUNPOUR ET AL., 1993]. TATE montre qu'un groupe d'auditeurs habitant la Floride, sans connaissances linguistiques, a reconnu respectivement 30 et 37,5 % de locuteurs originaires du sud des États-Unis dans deux groupes de locuteurs imitant

l'accent du sud des États-Unis, un groupe de locuteurs non entraînés et un groupe d'acteurs [TATE, 1979 IN : NOLAN, 1983]. Ce résultat peut être considéré comme élevé, étant donné la complexité des règles phonologiques qui gouvernent le système sonore complet d'un accent particulier.

4.3.3.3.3. Influence de la présence de voix auditivement proches (Tableau IV.5.)

La présence d'une paire de jumeaux univitellins dans un ensemble de douze locuteurs a provoqué 96 % de fausse acceptation de la part des auditeurs dans l'expérience de ROSENBERG [ROSENBERG, 1973]. D'autre part, HOMAYOUNPOUR ET CHOLLET montrent que les jumeaux ont des voix plus difficilement différenciables que des personnes ayant des voix similaires, mais sans lien de fraternité [HOMAYOUNPOUR ET CHOLLET, 1995]. D'autre part, l'aptitude des auditeurs familiers des jumeaux est plus grande que celle des auditeurs non familiers dans la tâche de discrimination de locuteurs. Le très faible taux d'acceptation correcte, mis en évidence par ROSENBERG, n'est pas confirmé par HOMAYOUNPOUR ET CHOLLET, probablement parce que les auditeurs de la seconde expérience étaient informés de la présence de paires de locuteurs jumeaux, alors que les auditeurs de la première ne l'étaient pas.

Auteurs	Ensemble de référence des locuteurs	Caractéristiques du message	Caractéristiques du canal de transmission	Auditeurs	Tâche et mesure de performance
ROSENBERG (1973)	A : 8 locuteurs imités par 4 imitateurs B : Une paire de jumeaux	Une phrase	Haute qualité	2 groupes de 5 femmes non familières	Discrimination : A : 60 % à 96 % B : 4 %
HOMAYOUNPOUR ET AL. (1993)	11 locuteurs et 13 locutrices de 25 à 50 ans	46 paires de phrases de 4 s imitées	Qualité téléphonique	20 auditeurs non familiers	Discrimination 68 % (faux rejet : 6 % ; fausse acceptation : 26 %)
HOMAYOUNPOUR ET CHOLLET (1995)	4 paires de jumeaux et 5 paires de jumelles univitellins 27 autres personnes incluant 4 non-jumeaux	46 paires de phrases de 6 s A : Aucune paire de phrases ne provient de jumeaux B : Toutes les paires de phrases proviennent de jumeaux	Qualité téléphonique	1 : Auditeurs non familiers 2 : Auditeurs familiers	Discrimination : A1 : 68 % (faux rejet : 17 % ; fausse acceptation : 15 %) A2 : 69 % (faux rejet : 17 % ; fausse acceptation : 14 %) B1 : 48 % (faux rejet : 29 % ; fausse acceptation : 23 %) B2 : 64 % (faux rejet : 20 % ; fausse acceptation : 16 %)

Tableau IV.5. Influence de la présence de voix auditivement proches

4.3.3.3.4. Détermination de caractéristiques physiques générales du locuteur (*profiling*)

Les premières expériences contrôlées s'attachent à mesurer l'aptitude de l'auditeur à déterminer les types physique et psychologique du locuteur [SAPIR, 1927 ; TAYLOR, 1933 ; HERZOG, 1933 ; ALLPORT ET CANTRIL, 1934 ; KAISER 1939-1944]. Selon plusieurs auteurs de cette époque, il est même possible de lui assigner un des trois types physiques décrits par KRETSCHMER [KRETSCHMER, 1922 ; BONAVENTURA, 1935 ; FAY ET MIDDLETON, 1940 ; MOSES, 1941]. Cependant MCGEHEE trouve que les performances des auditeurs sont faibles quant à l'évaluation de l'âge, de la taille, du poids et des caractéristiques de la personnalité des locuteurs [MCGEHEE, 1944].

Néanmoins, plusieurs études ont suggéré qu'une évaluation du poids et de la taille du locuteur pouvait être réalisée par des auditeurs profanes, à partir de la fréquence fondamentale [LASS ET AL., 1978, LASS ET AL., 1980A]. Cependant, ni la suppression de la fréquence fondamentale, ni le filtrage passe-bas ou passe-haut à 255 Hz [LASS ET AL. ; 1980B], ni même la suppression de la plage de fréquence des deux premiers formants [GUNTER ET MANNING, 1982], n'affecte le jugement des auditeurs de façon significative. Une mesure directe de la corrélation entre taille, poids et fréquence fondamentale montre que l'information sur le poids et la taille n'est pas localisée dans ce seul paramètre acoustique, mais contenue dans le signal de parole tout entier [KÜNZEL, 1989].

4.3.3.3.5. Détermination de la race du locuteur

La capacité de détermination de la race par des auditeurs profanes, à partir d'un signal de parole, a aussi été étudiée [MERTZ ET KIMMEL, 1978 ; LASS ET AL., 1980C]. Le critère de la couleur de la peau a été retenu comme critère de distinction entre les différents groupes de locuteurs. Or, sur le plan biologique, aucun critère précis ne permet de distinguer un groupe humain et la plupart des caractéristiques physiques varient de manière progressive et indépendante parmi les personnes [LANGANEY, 1992]. D'une part, dans une même population, les différences de couleur de peau peuvent être grandes et, d'autre part, une correspondance de la couleur de la peau n'implique aucune ressemblance de la forme et du fonctionnement des organes participant à la production de la parole.

Les ressemblances perçues dans la parole et attribuées à la race sont plutôt à rechercher dans des origines ethniques, culturelles et sociales communes. Sur le plan anatomique, certains petits muscles participant à la phonation comme les muscles thyro-épiglottiques inférieurs, thyromembraneux et crico-thyroïdiens, ou participant à l'articulation comme le muscle risorius, qui rétracte les lèvres pour sourire, ne sont pas présents chez toutes les ethnies, ou sont présents sous des formes différentes (Tableau IV.6.) [MERKEL, 1902 ; CATFORD, 1977]. Les résultats de LASS montrent d'ailleurs que la capacité à déterminer si le locuteur est noir ou blanc varie entre 60 et 70 %, résultat s'approchant d'une décision prise au hasard [LASS ET AL., 1980C].

Population	Présence du muscle risorius	Présence des muscles thyro-épiglottiques inférieurs et thyromembraneux	Muscles crico-thyroïdiens		
			Un seul muscle	Deux muscles solitaires au centre	Deux muscles indépendants
Européenne	85 % (Allemands)	75 à 80 %	0 %	10 %	90 %
Asiatique	20 % (Japonais)	80 à 100 % (Chinois)	8 % (Japonais)	34 % (Japonais)	57 % (Japonais)

Tableau IV.6. Particularités anatomiques de muscles participant à la phonation, selon la race [CATFORD, 1977]

4.3.3.3.6. Détermination du genre du locuteur (Tableau IV.7.)

COLEMAN montre que la détermination du sexe par des auditeurs est corrélée à 94% à la hauteur de la fréquence fondamentale et, dans une moindre mesure, à la résonance du tractus vocal (59%), estimée par la mesure des trois premiers formants [COLEMAN, 1976]. LASS confirme cette corrélation en montrant la dégradation des performances entre la détermination du sexe sur la base de voyelles voisées /i/, /e/, /œ/, /a/, /o/ et /u/ et des mêmes voyelles, chuchotées [LASS

ET AL., 1976]. La dégradation du message par un filtrage passe-bas ou passe-haut à 255 Hz, n'affecte par contre que peu la performance auditive de détermination du sexe du locuteur ; l'identification des femmes est toujours plus facile que celle des hommes [LASS ET AL., 1980B].

La suppression de l'information personnelle concernant la fréquence fondamentale, par l'utilisation d'un larynx artificiel générant une onde de 120 Hz ou de 240 Hz, montre que la détermination correcte du sexe sur la base de la résonance du tractus vocal est de 67 % pour les hommes et seulement de 30 % pour les femmes [COLEMAN, 1976]. La discrimination de paires de locuteurs reste cependant possible dans plus de 90 % des cas [COLEMAN, 1973].

Auteurs	Ensemble de référence des locuteurs	Caractéristiques du message	Caractéristiques du canal de transmission	Auditeurs	Tâche et mesure de performance
COLEMAN (1973)	20 locuteurs 20 locutrices	Lecture du « <i>Rainbow Passage</i> »	Haute qualité	17 auditeurs adultes et jeunes	Corrélation entre la détermination du sexe et F_0 : 0.94 et F1 , F2 et F3 : 0.59
LASS ET AL. (1976)	10 locuteurs 10 locutrices	A : /i/ B : /ε/ C : /æ/ D : /ɑ/ E : /o/ F : /u/ 1 : voix normale 2 : voix chuchotée	Haute qualité	15 auditeurs	Détermination ♀ / ♂ : A1 : 96 % A2 : 74 % B1 : 96 % B2 : 80 % C1 : 98 % C2 : 76 % D1 : 98 % D2 : 79 % E1 : 94 % E2 : 74 % F1 : 94 % F2 : 68 %
LASS ET AL. (1980C)	A : 10 locuteurs B : 10 locutrices	4 phrases	1 : Haute qualité 2 : Filtrage passe-bas 3 : Filtrage passe-haut	28 auditeurs	Détermination ♀ / ♂ : A1 : 96 % B1 : 99 % A2 : 92 % B2 : 98 % A3 : 95 % B3 : 96 %
COLEMAN (1976)	A : 5 locuteurs choisis : (F1, F2, F3 bas) B : 5 locutrices choisies : (F1, F2, F3 haut) 1 : F_0 artificielle : 120 Hz 2 : F_0 artificielle : 240 Hz	Lecture du « <i>Rainbow Passage</i> »	Haute qualité	18 auditrices et 7 auditeurs adultes et jeunes	Détermination ♀ / ♂ : A1 : 100 % A2 : 67 % B1 : 30 % B2 : 96 %
INGEMAN (1968)	14 locuteurs phonétiquement entraînés	A : /h/ B : /ʃ/ C : /s/ D : /ʒ/ E : /x/ F : /θ/ G : /c/ H : /ϕ/ I : /f/	Haute qualité	5 auditeurs 5 auditrices	Détermination ♀ / ♂ : A : 91 % B : 77 % C : 75 % D : 73 % E : 67 % F : 61 % G : 60 % H : 55 % I : 54 %

Tableau IV.7. Détermination du genre du locuteur

Les fricatives sourdes, surtout /h/, /ʃ/ et /s/ prononcées de façon isolée, contribuent aussi à la détermination du sexe du locuteur, malgré l'absence d'information concernant la fréquence

fondamentale [INGEMAN, 1968 ; SCHWARTZ, 1968]. SCHWARTZ ET RINE soulignent cependant que cette détermination est plus facile à partir de mots isolés qu'à partir d'un texte continu [SCHWARTZ ET RINE, 1968].

4.3.3.3.7. Détermination de l'âge du locuteur (Tableau IV.8.)

L'examen de plusieurs centaines de locuteurs et locutrices âgés de six à nonante ans montre que pendant l'enfance et l'adolescence la tessiture s'élargit et la hauteur moyenne de la fréquence fondamentale s'abaisse, avant de se stabiliser à l'âge adulte. Au-delà de la soixantaine, la tessiture rétrécit, principalement à cause du déplacement de la limite inférieure. Simultanément, la hauteur moyenne de la fréquence fondamentale augmente chez l'homme, alors qu'elle diminue chez la femme [BÖHME ET HECKER, 1970]. Pour les locuteurs, cette constatation est confirmée par HOLLIEN ET SHIPP, qui mettent en évidence une légère diminution de la hauteur moyenne de la fréquence fondamentale de 120 Hz à 112 Hz entre 20 et 40 ans, aussi relevée par SUZUKI, et une augmentation linéaire de 107 Hz à 146 Hz entre 40 et 90 ans [HOLLIEN ET SHIPP, 1972 ; SUZUKI ET AL., 1994]. HOLLIEN ET SHIPP relèvent cependant l'existence de larges différences individuelles.

Auteurs	Ensemble de référence des locuteurs	Caractéristiques du message	Caractéristiques du canal de transmission	Auditeurs	Tâche et mesure de performance
SHIPP ET HOLLIEN (1969)	7 x 25 locuteurs de 20 à 90 ans regroupés par décennies	3 ^{ème} phrase du « <i>Rainbow Passage</i> »	Haute qualité	25 auditeurs adultes, jeunes	Estimation directe de l'âge Corrélation entre âge chronologique et âge perçu : 0.88
HORII ET RYAN (1981)	57 locuteurs de 40 à 80 ans A : 29 locuteur dont la voix correspond à l'âge perçu B : 28 locuteurs dont la voix ne correspond pas à l'âge perçu	1 ^{er} paragraphe du « <i>Rainbow Passage</i> »	Haute qualité	20 auditeurs	Estimation directe de l'âge Corrélation entre âge chronologique et âge perçu : A : 0.84 B : 0.67 A + B : 0.76
NEIMAN ET APPLGATE (1990)	A : 18 locuteurs B : 18 locutrices 3 par tranche d'âge : 1 : 20 - 25 2 : 30 - 35 3 : 40 - 45 4 : 50 - 55 5 : 60 - 65 6 : 70 - 75	3 premières phrases du « <i>Rainbow Passage</i> »	Haute qualité	Groupe d'auditeurs	Estimation correcte de la décennie : A1 : 61 % B1 : 91 % A2 : 84% B2 : 85 % A3 : 81 % B3 : 74 % A4 : 90 % B4 : 86 % A5 : 81 % B5 : 82 % A6 : 69 % B6 : 80 % Total 80.25 % ± 11.05
BRAUN ET RIETVELD (1995)	40 locuteurs de 27 à 59 ans (m : 41.05 ans) A : 20 non-fumeurs B : 20 fumeurs de 10 à 40 ans	Lecture du texte « <i>The North Wind and the Sun</i> » en allemand (45 s)	Haute qualité	19 auditeurs de 20 à 32 ans	Estimation directe de l'âge A : 37.40 ans B : 43.79 ans A + B : 40.59 ans

Tableau IV.8. Détermination de l'âge du locuteur

À l'aide d'auditeurs entraînés, les principaux indicateurs de l'âge du locuteur ont pu être mis en évidence : la hauteur de la fréquence fondamentale, la vitesse d'élocution [HARTMAN ET DANHAUER, 1976], le tremblement de la voix, la tension du larynx, la perte d'air, l'imprécision des consonnes et la vitesse de l'articulation [RYAN ET BURK, 1972].

« Lorsque des locuteurs lisent un même passage, leur âge peut habituellement être évalué dans une tranche de dix ans » [ALLPORT, 1963]. Cette constatation a été illustrée par de nombreuses expériences, après que PTACEK ET SANDER eurent mis en évidence des différences physiologiques lors de la phonation dans un groupe de locuteurs de moins de quarante ans, par rapport à un groupe de locuteurs de plus de soixante-cinq ans [PTACEK ET SANDER, 1966]. SHIPP et HOLLIEN montrent la capacité des auditeurs à classer des locuteurs comme jeunes, adultes ou vieux et à déterminer leur décennie [SHIPP ET HOLLIEN, 1969]. Lors d'une estimation directe de l'âge, ils établissent une corrélation de 0.88 entre l'âge chronologique (CA) des locuteurs et l'âge perçu (PA). Ces résultats ont été confirmés notamment par RYAN et BURK, avec une corrélation de 0.74, par HORII ET RYAN, avec une corrélation de 0.76, par NEIMAN et APPLGATE avec une corrélation de 0.88 et par BRAUN, avec une corrélation de 0.68 [RYAN ET BURK, 1974 ; HORII ET RYAN, 1981 ; NEIMAN ET APPLGATE, 1990 ; BRAUN, 1996]. Seule exception, l'étude de RAMIG qui ne met en évidence qu'une corrélation de 0.17 [RAMIG ET AL., 1985 IN : BRAUN, 1996].

La perception de l'âge des locuteurs dépend aussi de l'âge des auditeurs [HUNTLEY ET AL. ; 1987], de la différence d'âge entre locuteurs et auditeurs [SHIPP ET HOLLIEN, 1969] et du sexe des auditeurs [HARTMANN, 1979]. L'état physiologique du locuteur influence aussi la perception de son âge par la voix [RAMIG ET RINGEL, 1983]. La voix des locuteurs en bonne santé est perçue comme plus jeune que les autres [RINGEL ET CHODZKO-ZAJKO, 1987]. La consommation de tabac, sous forme de fumée, modifie l'état physiologique et histologique des organes participant à la phonation et les fumeurs sont perçus comme étant plus âgés que les non-fumeurs [BRAUN ET RIETVELD, 1995].

4.3.3.3.8. Influence du temps écoulé entre l'écoute du message et l'audition de comparaison (Tableau IV.9.)

Dans l'affaire *State v Hauptmann*³¹, le colonel Charles Lindbergh prétendit identifier la voix de l'accusé comme étant celle du ravisseur de son fils kidnappé presque trois ans auparavant. Bien que son témoignage fut accepté en cour, la défense déclara qu'une telle identification ne pouvait pas avoir valeur de preuve.

Suite à cette affaire, MCGEHEE a montré que la fiabilité de l'identification diminue rapidement lorsque l'intervalle de temps entre les sessions est supérieur à deux semaines [MCGEHEE, 1937 ; MCGEHEE, 1944]. Certains aspects de ces données sont cependant en opposition, dans le sens que l'étude de 1937 montre que le taux d'identification passe de 68 % après deux semaines, à 51 % après trois semaines et 35 % après trois mois, alors qu'en 1944, ce taux passe de 48 % à 47 % et 45 % respectivement après deux, quatre et huit semaines. Ces résultats permettent

³¹ [State v Hauptmann, Atlantic Rep., 1935, 180, 809 - 829]

tout de même de circonscrire une estimation de la limite temporelle de la mémoire à long terme, impliquée dans la tâche d'identification de locuteurs.

Lorsque la voix est mémorisée incidemment par l'auditeur, le taux d'identification d'un locuteur ou d'une locutrice dans un groupe de onze personnes est d'environ 45 % lorsque le délai entre la première écoute et la confrontation est de quelques heures. Lorsque ce délai se situe entre un jour et deux semaines, le taux d'identification chute à 20 %. Après un délai de trois semaines, le résultat n'est plus que de 9 %, ce qui équivaut à un choix effectué au hasard [CLIFFORD ET DENOT, 1980 *IN* : CLIFFORD, 1980]. Ces résultats montrent que les performances de l'humain sont faibles, lorsque la durée entre la première écoute et la suivante augmente, mais la décroissance monotone observée par [MCGEHEE, 1937 ; MCGEHEE, 1944] n'est pas confirmée.

Auteurs	Ensemble de référence des locuteurs	Caractéristiques du message	Caractéristiques du canal de transmission	Auditeurs	Tâche et mesure de performance
MCGEHEE (1937)	5 locuteurs	Lecture d'un paragraphe de texte à vitesse normale et altérée	Haute qualité	Groupe d'auditeurs non familiers	Identification après : 1 jour : 83 % 3 jours : 81 % 1 sem. : 81 % 2 sem. : 69 % 1 mois : 57 % 5 mois : 13 %
CLIFFORD ET DENOT (1980) <i>IN</i> : CLIFFORD (1980)	11 locuteurs et 11 locutrices	Une phrase	Haute qualité	A : 70 auditeurs B : 44 auditeurs	Identification après : A : 10 min. : 56 % A : 40 min. : 45 % A : 100 min. : 40 % A : 130 min. : 44 % B : 10 min. : 41 % B : 1 jour : 20 % B : 1 sem. : 23 % B : 2 sem. : 19 %

Tableau IV.9. Influence du temps écoulé entre l'écoute du message et l'audition de comparaison

4.3.3.3.9. Influence de la durée du message (Tableau IV.10.)

Les performances de l'humain dépendent aussi de la quantité et de la qualité du message [STEVENS *ET AL.*, 1968]. Si le taux d'identification entre huit locuteurs familiers est de 45 % après 0,2 s, il est déjà de 98 % après 2 s [POLLACK *ET AL.*, 1954]. Sur la base d'échantillons de parole de 25 ms, le taux d'identification de voix familières est déjà supérieur au résultat issu d'un choix effectué au hasard [COMPTON, 1963]. La progression de ce taux est forte jusqu'à 1,2 s et diminue par la suite [BOLT *ET AL.* ; 1970].

Plus encore que la durée, la richesse phonétique est déterminante ; de 56 % sur la base d'un seul phonème, le taux d'identification de seize locuteurs par dix auditeurs familiers progresse à 98 %, sur la base d'une phrase de plus de quinze phonèmes [BRICKER ET PRUZANSKY, 1966]. STEVENS *ET AL.* montrent aussi une augmentation appréciable du taux de reconnaissance sur la base d'échantillons d'une, de deux ou de trois syllabes [STEVENS *ET AL.*, 1968].

Avec des auditeurs adultes, CLIFFORD montre que l'augmentation du taux d'identification à partir d'un échantillon d'une, de deux et de quatre phrases n'est pas significatif, mais met en évidence le fait que les voix de femmes sont significativement mieux reconnues que les voix d'hommes, 85 % contre 70 %, confirmant ainsi l'observation de MCGEHEE [MCGEHEE, 1944 ; CLIFFORD, 1980]. La même expérience, effectuée avec des auditeurs âgés de 12 à 16 ans à partir de messages plus courts, d'une demi, d'une et de deux phrases, montre une différence significative des résultats en fonction de la durée du message. Les performances en moyenne plus faibles des jeunes auditeurs laissent à penser que leurs capacités d'identification sont moindres que celles des adultes [CLIFFORD, 1980]. L'évolution des performances de discrimination de locuteurs ayant des voix auditivement similaires en fonction de la durée du message a été mise en évidence par ROSE ET DUNCAN, qui montrent par la même occasion les différences de performances entre les auditeurs familiers et non familiers [ROSE ET DUNCAN, 1995].

Auteurs	Ensemble de référence des locuteurs	Caractéristiques du message	Caractéristiques du canal de transmission	Auditeurs	Tâche et mesure de performance
POLLACK ET AL. (1954)	de 2 à 8 locuteurs	Parole spontanée (~3.5 syl./s)	Haute qualité	7	Identification : 0,2 s : 45 % 0,43 s : 78 % 0,65 s : 83 % 1,15 s : 95 % 2 s : 98 %
BRICKER ET PRUZANSKY (1966)	10 locuteurs familiers	A : Phrase (> 15 ph.) B : Disyllabes (4 ph.) C : Monosyllabes (3,2 ph.) D : Consonne-voyelle (2 ph.) E : Voyelles (1 ph.)	Haute qualité	16	Identification : A : 98 % B : 87 % C : 81 % D : 63 % E : 56 %
STEVENS ET AL. (1968)	8 locuteurs	A : Une phrase B : Une syllabe	Haute qualité	6	Identification : A : 92 % B : 88 %
CLIFFORD (1980)	6 locuteurs	A : demi-phrase B : Une phrase C : Deux phrases D : Quatre phrases	Haute qualité	1 : 134 auditeurs adultes 2 : 132 auditeurs de 12 à 16 ans	Identification : A2 : 36 % B1 : 75 % B2 : 41 % C1 : 77 % C2 : 49 % D1 : 82%
ROSE ET DUNCAN (1995)	6 locuteurs ayant des voix similaires : A : 4 familiers B : 2 non familiers	1 : 45 s de parole spontanée 2 : Un seul mot	Haute qualité	10	Discrimination : A1 : 85 % A2 : 74 % B1 : 67 % B2 : 45 %

Tableau IV.10. Influence de la durée du message

4.3.3.3.10. Influence des caractéristiques du canal de transmission (Tableau IV.11.)

La diminution de la bande passante du canal de transmission affecte aussi les performances de reconnaissance de locuteurs. L'effet de filtres passe-haut et passe-bas est symétrique et les courbes de dégradation des performances sont centrées en 1500 Hz [POLLACK *ET AL.*, 1954]. À partir d'échantillons de la voyelle [i] de 25 ms à une seconde et demie, COMPTON montre que, pour un taux d'identification donné, la durée de l'échantillon doit augmenter si la largeur de bande du canal de transmission diminue [COMPTON, 1963]. De même, le taux de reconnaissance de locuteurs, à partir de signaux transmis par téléphone ou ayant traversé un codeur LPC, est plus faible qu'à partir de signaux exempts de ces distorsions [PAPAMICHALIS ET DODDINGTON, 1984].

Auteurs	Ensemble de référence des locuteurs	Caractéristiques du message	Caractéristiques du canal de transmission	Auditeurs	Tâche et mesure de performance
POLLACK <i>ET AL.</i> (1954)	8 locuteurs	Parole spontanée (~3.5 syl./s)	A : Filtrage passe-bas B : Filtrage passe-haut 1 : 100 Hz 2 : 250 Hz 3 : 500 Hz 4 : 1 KHz 5 : 2 KHz 6 : 5 KHz	7	Identification : A1 : 53 % B1 : 84 % A2 : 54 % B2 : 83 % A3 : 61 % B3 : 81 % A4 : 70 % B4 : 78 % A5 : 79 % B5 : 72 % A6 : 84 % B6 : 55 %
MCGONEGAL <i>ET AL.</i> (1978)	A : 8 locuteurs et 8 locutrices enregistrés 10 fois sur plusieurs semaines B : 62 imposteurs : 31 ♀ et 31 ♂ enregistrés une fois	♂ : « <i>We were away a year ago</i> » ♀ : « <i>I know when my lawyer is due</i> »	1 : Filtre 100 Hz -2.6 KHz 2 : Codage ADPCM et filtre 1 3 : Codage LPC et filtre 1	Auditeurs non familiers et inexperts	Discrimination locuteur/imposteur : A1B1 ♀ : 89 % ♂ : 85 % A2B2 ♀ : 86 % ♂ : 83 % A3B3 : ♀ : 80 % ♂ : 81 % A2B1 ♀ : 81 % ♂ : 80 % A3B1 ♀ : 79 % ♂ : 86 % A3B2 ♀ : 85 % ♂ : 82 %
SCHMIDT-NIELSEN ET STERN (1985)	A : 19 locuteurs familiers B : 5 locuteurs non familiers	Message de haute qualité : 20 s à 40 s Message codé LPC : 27 s à 81 s	1 : Haute qualité 2 : Codage LPC : 2.4 Kbit/s	24 auditeurs	Identification : A1 : 90 % A2 : 71 % A1 + B1 : 88 % A2 + B2 : 69 %

Tableau IV.10. Caractéristiques du canal de transmission

L'étude de MCGONEGAL montre qu'en cas de transmission numérique le système de codage utilisé n'a que très peu d'influence sur les performances de discrimination, lorsque celui-ci est homogène pour tous les échantillons. Cependant, comme la qualité de la voix résultant d'un codage LPC ou ADPCM ³² est extrêmement différente, les performances sont significativement altérées lorsque des systèmes différents sont utilisés pour le codage de l'indice et des enregistrements de comparaison [MCGONEGAL *ET AL.*, 1978].

³² *supra* : 2.3.3.2. Réseau téléphonique public commuté (RTPC)

³³ *supra* : 2.3.3.4. Communications sécurisées et communications par satellite

Issus de la recherche militaire, les travaux de SCHMIDT-NIELSEN et STERN font appel à un système de codage inconnu dans les systèmes de communication civils, le codage LPC à 2,4 Kbits s⁻¹, dont la qualité de transmission est nettement inférieure à ces derniers : MOS : 2.5, DRT : 88³³ [SCHMIDT-NIELSEN ET STERN, 1985]. Ils mettent cependant en évidence que certains locuteurs, bien reconnus lorsque le signal est de haute qualité, ne sont que faiblement reconnus lorsque le signal est codé ; ils ne mesurent d'ailleurs qu'une corrélation de 0.66 entre le taux d'identification d'échantillons de haute qualité et codés.

4.3.4. Limites de la procédure de reconnaissance par des profanes

L'investigation expérimentale de l'aptitude de la personne inexperte dans la tâche de reconnaissance de locuteurs montre qu'un grand nombre de paramètres conditionnent ses performances. Ce sont notamment la nature de la tâche demandée, la taille et l'homogénéité de l'ensemble de référence des locuteurs, l'âge et le sexe des locuteurs et des auditeurs, la qualité et la quantité de parole entendue initialement, le délai entre l'écoute initiale de la voix inconnue et la procédure d'écoute et d'identification par les victimes ou les témoins, le déguisement de la voix durant l'écoute initiale, les différents moyens techniques de transmission et d'enregistrement utilisés et la présence ou l'absence d'un témoignage visuel concordant [CLIFFORD *ET AL.* ; 1981].

Même si tous les paramètres sont soigneusement choisis en vue d'augmenter la validité d'une telle procédure, la grande variabilité des performances individuelles des auditeurs limite le résultat d'une telle investigation à une valeur indicative, dont l'incertitude est à rapprocher de celle du témoignage [HOLLIEN *ET AL.* 1982 ; HAMMERSLEY ET READ, 1983].

Le résultat obtenu lorsqu'un échantillon de parole inconnue est diffusé à grande échelle, par exemple à la radio, semble plus intéressant. En effet, dans cette circonstance les auditeurs familiers du locuteur inconnu, touchés par une telle diffusion, peuvent procéder à une tâche d'identification basée sur une référence mnémonique étendue et non pas à une tâche de discrimination ou à une tâche d'identification basée sur une référence limitée à un stimulus, comme c'est le cas pour les victimes ou les témoins.

4.4. Procédure de reconnaissance par des experts

L'approche actuelle pratiquée par les phonéticiens combine l'approche phonétique auditive perceptive et les techniques de phonétique acoustique [KÜNZEL, 1987 ; BRAUN, 1995]. L'analyse linguistique forensique ne fait pas partie de cette investigation, mais elle constitue une méthode d'analyse de l'échantillon de parole inconnue supplémentaire et indépendante [KNIFFKA, 1990 *IN* : NOLAN, 1990].

4.4.1. L'approche auditive perceptive

Cette approche est basée sur l'analyse auditive de paramètres comme la qualité de la voix, les défauts de prononciation et d'élocution, la vitesse d'élocution, l'intonation et le rythme, ainsi que les analyses d'ordre plus élevé : lexicale, syntaxique, sémantique et idiomatique, en vue de la caractérisation du dialecte et de l'accent régional. Elle peut aussi considérer des paramètres paralinguistiques, comme le cycle inspiration-expiration ou la durée des silences [BRAUN, 1995].

4.4.1.1. Caractéristiques segmentales

Au début du travail de comparaison des échantillons de parole, une pratique consensuelle des phonéticiens consiste à transcrire phonétiquement les échantillons inconnus et de comparaison. Ils se basent sur les symboles de l'Association Phonétique Internationale (IPA) pour retranscrire les moindres détails de la prononciation des voyelles et des consonnes, souvent en conjonction avec l'utilisation de symboles supplémentaires, propres à chaque expert. Une attention particulière est portée à chaque caractéristique que l'analyste considère comme idiosyncrasique. Ce choix dépend évidemment de la connaissance que l'analyste a des accents régionaux et sociaux concernés, dans le sens que l'habilité à identifier des déviations d'une norme phonologique présuppose une compréhension ou une familiarité avec cette norme [FRENCH, 1994].

Une compréhension des normes phonologiques est un facteur qui doit aussi être pris en compte lorsque l'expert analyse des échantillons dont la langue lui est étrangère. Le code de procédure édicté par l'*International Association for Forensic Phonetics* (IAFP)³⁴ suggère à ses membres d'approcher avec la plus grande prudence l'analyse forensique d'échantillons de parole énoncés dans une autre langue que leur langue maternelle [NOLAN, 1992]. En pratique, le choix des caractéristiques devrait être effectué avec l'assistance d'une personne de la langue concernée, ayant des connaissances phonétiques et linguistiques, ou une personne ayant des connaissances phonologiques étendues de la langue analysée [FRENCH, 1994].

4.4.1.2. Caractéristiques suprasegmentales ou prosodiques

4.4.1.2.1. Fréquence fondamentale moyenne, rythme et aisance d'expression

Les examens auditifs de la fréquence fondamentale peuvent aussi inclure une analyse de l'intonation et une première évaluation de la hauteur moyenne de la fréquence fondamentale.

³⁴ *infra* : Annexe V. Code de procédure de l'IAFP

Aucune relation statistiquement significative n'a été trouvée entre la hauteur de la fréquence fondamentale, la vitesse d'élocution et son intensité [KÜNZEL *ET AL.*, 1995]. Il n'a pas été démontré non plus que, dans une langue donnée, la fréquence fondamentale moyenne varie avec des paramètres régionaux ou sociaux ; c'est par contre le cas pour l'intonation. Des déviations par rapport à la norme de l'accent peuvent être notées et sélectionnées pour un examen acoustique ultérieur. La fréquence fondamentale est déterminée par une analyse acoustique d'échantillons comparables, sélectionnés sur la base d'une impression auditive préliminaire [FRENCH, 1994].

Le rythme de la parole et l'aisance d'expression sont aussi étudiés, ainsi que les figures d'élision et d'assimilation. Ces aspects de la parole contiennent un potentiel d'identification individuelle, étant donné les divergences qui peuvent apparaître entre locuteurs d'un même milieu social et régional [FRENCH, 1994].

4.4.1.2.2. Timbre ou qualité de la voix

Le timbre est la combinaison des résonances obtenues par la modulation du signal acoustique lors de l'articulation et qui modifient sa composition spectrale. Comme il s'agit certainement de l'élément le plus subjectif, mais aussi le plus représentatif d'une voix donnée, il est beaucoup plus ouvert à une analyse auditive perceptive qu'à une analyse phonétique acoustique. Les impressions auditives du timbre peuvent être notées soit de façon informelle, soit à l'aide d'un système d'évaluation formalisé [LAVER, 1980]. Un enregistrement composé d'une alternance d'extraits d'échantillons inconnus et de comparaison peut permettre de comparer les timbres. Bien que de tels montages soient le plus souvent réalisés au profit de l'expert, ils sont parfois produits au tribunal pour démontrer la méthode de comparaison des timbres utilisée par le phonéticien ou pour illustrer les différences ou les ressemblances. Certains phonéticiens se sont prononcés contre cet usage, car il peut détourner l'attention de la cour qui substituera, peut-être de manière inconsciente, son propre jugement à celui de l'expert [FRENCH, 1994].

4.4.1.3. Approche systématique

Par analogie à la méthode graphoscopique de LOCARD, et dans un but de classification des échantillons de parole inconnue, FÄHRMANN a proposé une systématisation de cette approche, articulée en trois éléments [LOCARD, 1959 ; FÄHRMANN, 1966A ; FÄHRMANN, 1966B] :

- A. L'analyse de la structure du texte, qui comprend l'étude du contenu de la conversation, la prédominance de l'accent, les défauts du langage et l'habileté du langage.
- B. L'analyse de la voix et du langage, qui intègre l'impression d'ensemble (du général au particulier), l'étude de la forme composée de l'analyse de la construction de la phrase, du choix des mots, de la qualité des mots, de la diction, du texte, du style de langage et de sa dynamique.
- C. L'analyse des particularités de la voix, qui englobe l'analyse de la hauteur du son, de sa force, de la plénitude de la voix, du timbre, du tempo, du rythme des mots et des phrases, du degré d'excitation du locuteur, de la mélodie du langage, de

l'articulation et de l'accentuation composée de la mélodie du langage, de l'accent dynamique, de la résolution, de l'allure des pauses et de la modulation du timbre.

4.4.1.4. Limites de l'approche phonétique auditive perceptive

L'émergence de cette pratique, principalement au Royaume-Uni, semble plutôt muer par l'augmentation de la demande et non la conséquence d'un développement scientifique. Les principes sur lesquels se base cette méthode n'ont fait l'objet que de peu de publications et il n'est pas évident que les pratiques des phonéticiens cités au tribunal soient suffisamment unifiées pour que l'on puisse les considérer comme une seule et même méthode. Il apparaît cependant que les praticiens se basent sur une combinaison de l'impression auditive générale, de la qualité et de la hauteur de la voix et sur la comparaison d'une variété de segments phonétiques présents dans les échantillons. L'utilité de la seconde partie de cette méthode repose sur la supposition que la conjonction d'une sélection suffisante de variables segmentales aboutit non seulement à la catégorisation d'un accent, mais à la description d'un idiolecte propre à l'individu [NOLAN, 1990].

Incontestablement cette seconde partie, essentiellement dialectologique, permet d'éliminer une suspicion nourrie à l'encontre de certains locuteurs suspects, par la mise en évidence de différences d'accent conséquentes entre l'échantillon inconnu et celui de comparaison. De plus, il semble improbable qu'un locuteur, dans une volonté de déguisement, conjugue cohérence et exhaustivité dans l'accent qu'il adopte [NOLAN, 1990].

Par contre, l'utilisation de cette méthode dans une procédure d'identification nécessite soit que le processus de caractérisation aboutisse à une spécificité plutôt qu'à la sélection d'un groupe, soit que l'analyse auditive seule, incluant l'analyse segmentale et les observations globales de la qualité et de la hauteur de la voix, permette l'identification. Or, aucune expérience contrôlée n'a été menée pour déterminer s'il existe des individus dont la qualité et la hauteur de la voix ainsi que l'accent sont si proches qu'ils ne peuvent être discriminés par l'analyse auditive des variables segmentales. Le nombre des variables considérées dépend d'ailleurs aussi de la taille de l'échantillon de parole inconnue.

Si l'hypothèse théorique qu'aucun être humain ne prononce d'énoncé de parole de manière identique demeure, l'absence d'une démonstration à grande échelle empêche l'extension de l'observation de cette idiosyncrasie à un ensemble homogène de locuteurs. Dans ces conditions, le concept d'idiolecte semble de peu d'utilité dans la discrimination de locuteurs possédant un accent comparable et ne permet certainement pas d'aboutir à l'identification d'un locuteur « au delà du doute raisonnable ».

Tant qu'une recherche extensive ne démontrera pas l'utilité de l'idiolecte dans la reconnaissance de locuteurs, il faudra supposer que des échantillons, dont les propriétés phonétiques-linguistiques correspondent, peuvent provenir de locuteurs différents [NOLAN, 1990 ; NOLAN, 1991]. D'autre part, l'expérience de SHIRT montre que l'aptitude d'un groupe de phonéticiens à identifier un locuteur à partir d'échantillons de parole brefs sur la base de l'impression auditive générale, de la qualité et de la hauteur de la voix, est à peine supérieure à celle d'un groupe d'auditeurs inexperts [SHIRT, 1984 ; NOLAN, 1990].

Rares sont les phonéticiens qui, comme BALDWIN, considèrent l'approche auditive perceptive apte à identifier le locuteur à elle seule, mais nombreux sont ceux qui combattent l'argumentation développée par NOLAN [BALDWIN *IN* : BALDWIN ET FRENCH, 1990 ; NOLAN, 1990 ; NOLAN, 1991]. KÜNZEL et BRAUN notamment citent une étude de KÖSTER, qui montre une certaine supériorité des phonéticiens dans la tâche de reconnaissance auditive de locuteurs par rapport aux deux groupes d'auditeurs inexperts, 100% contre 94% et 89% [KÖSTER, 1987 ; KÜNZEL, 1994B ; BRAUN, 1995].

Par contre, l'approche qui combine la méthode auditive-perceptive traditionnelle et les techniques de phonétique acoustique rencontre l'adhésion de la plupart des phonéticiens [KÜNZEL, 1987 ; FRENCH *IN* : BALDWIN ET FRENCH, 1990 ; NOLAN, 1990 ; BRAUN, 1995].

4.4.2. L'approche phonétique acoustique

La méthode spectrographique permet la visualisation du signal de parole en trois dimensions, temps-fréquence-intensité, et ouvre la possibilité d'effectuer des mesures précises et rapides d'éléments sous-segmentaux et segmentaux du signal de parole dans ces trois dimensions. Elle joue un rôle dans l'approche d'ensemble, mais ne constitue pas le centre exclusif des examens acoustiques, comme c'est le cas de la reconnaissance de locuteurs par spectrogrammes. Les ressemblances et différences relevées par cette méthode de visualisation sont interprétées avec prudence et confrontées aux résultats d'analyse des échantillons par d'autres types d'investigation phonétique, à toute connaissance normative disponible et à l'expérience accumulée par l'analyste [FRENCH, 1994].

4.4.2.1. Caractéristiques segmentales fréquentielles

4.4.2.1.1. Fréquences formantiques

La production des voyelles orales est caractérisée par une excitation des cordes vocales, sans point d'articulation ni couplage nasal. Elle peut être modélisée par l'excitation d'un tube non uniforme par une pulsation quasi périodique. La réponse impulsionnelle du tube est caractérisée par ses fréquences de résonance et leurs harmoniques. Comme le tractus vocal n'est pas uniforme, les fréquences propres sont inégalement espacées en fréquence. Ces zones, où l'intensité des harmoniques est plus importante, sont appelées formants ou fréquences formantiques. La perception des voyelles est largement déterminée par leurs trois premiers formants, abrégés F_1 , F_2 et F_3 [DODDINGTON, 1970 ; FANT, 1973]. Si la fréquence fondamentale d'un locuteur augmente, alors qu'il conserve la même articulation, les indices d'harmoniques diminuent, alors que les formants ne changent pas [ORMEZZANO ET ROCH, 1991]. La mesure de la largeur de bande des formants, qui reste une opération difficile, a été réalisée notamment par [FURUI, 1989] : F_1 de 30 à 120Hz (moyenne : 50Hz), F_2 de 30 à 200 Hz (moyenne : 60Hz), F_3 de 40 à 300Hz (moyenne : 115Hz).

Voyelles françaises	1 ^{er} formant (F ₁) [Hz]	2 ^{ème} formant (F ₂) [Hz]	3 ^{ème} formant (F ₃) [Hz]
[i]	280	2300	2950
[e]	350	1950	2550
[ɛ]	450	1800	2470
[a]	660	1350	2380
[ɑ]	620	1150	2250
[ɔ]	480	1050	2250
[o]	360	780	2230
[u]	290	850	2270
[y]	290	1800	2140
[ø]	360	1450	2290
[œ]	490	1380	2270
[ə]	480	1400	2200

Tableau IV.11. Valeurs formantiques des voyelles orales du français [GROSJEAN, 1995]

Il existe une relation entre la forme du tractus vocal et l'enveloppe spectrale des voyelles [PERKELL *ET AL.*, 1986]. Les consonnes n'ont à l'origine pas de phase stationnaire, elles sont classées en voisées et non voisées et leurs caractéristiques dépendent largement des voyelles adjacentes, à cause du phénomène de coarticulation [FURUI, 1989]. Toutefois une dépendance certaine existe entre la forme du résonateur formé par les fosses nasales et les caractéristiques des consonnes nasales /n/ et /m/ [MELLA, 1992].

L'étude physiologique et acoustique de la valeur des trois premiers formants chez les hommes et les femmes montre que la longueur individuelle des cavités, et donc les valeurs des formants, peut changer de façon importante pour une catégorie d'âge et de sexe. Comme le larynx est placé plus bas chez les hommes, le pharynx est plus long. Ceci se traduit par des coefficients d'écart différents entre les formants, selon les voyelles et leur degré d'affiliation avec la partie pharyngale du conduit vocal. Chez les femmes, les valeurs des trois premiers formants sont, en moyenne, 18% plus élevées que chez les hommes. Pour une voyelle neutre, la variation de taille de la cavité est proportionnelle à la variation de fréquence des formants [FANT, 1973]. Pour les voyelles arrières, F₂ est corrélé à F₁, par contre, il est corrélé à F₃ pour les voyelles avant [PERKELL *ET AL.*, 1986].

Même si une dépendance au locuteur dans des voyelles isolées a été montrée pour les deux premiers formants, leur rôle est essentiellement phonétique car ils conditionnent la compréhension des voyelles [CALINSKI *ET AL.*, 1970]. Ce sont les formants d'ordre plus élevé qui conditionnent le plus la qualité de la voix du locuteur. Cependant, ils ont une étendue plus grande et une intensité plus faible que les deux premiers et, dans la qualité de parole téléphonique, ces formants d'ordre élevé manquent ou sont faiblement représentés [DODDINGTON, 1970, BALDWIN ET FRENCH, 1990].

La fiabilité des formants dépend de la localisation syntaxique et sémantique de la voyelle dans la phrase. En français, la meilleure fiabilité se retrouve dans les réalisations qui portent

l'accentuation linguistique. Cette constatation, aussi valable pour l'anglais, peut probablement être généralisée.

Les règles d'accentuation sont très différentes d'une langue à l'autre. En français l'accentuation est définie au niveau du syntagme et le rythme est celui du comptage. Chaque fois que la syntaxe s'arrête, la syllabe est accentuée ; le japonais est l'une des rares langues à partager ce trait avec le français. En anglais, l'accentuation est définie au niveau du mot. Elle se situe généralement sur la première syllabe des mots polysyllabiques. Les mots à étymologie latine suivent l'accentuation latine, qui se trouve sur l'avant-dernière syllabe si elle est longue ou sur l'antépénultième si l'avant-dernière est courte ; les autres mots suivent l'accentuation germanique et les mots très longs se comportent comme autant de mots courts [DELATTRE, 1965].

Par contre, les réalisations non accentuées et celles situées dans des mots grammaticaux, comme les auxiliaires ou les mots de liaison, sont moins robustes. Selon MELLA, les éléments les plus informatifs et les plus robustes sont, en français, les voyelles /e/, /œ/ et /ɔ/, et plus précisément F3 pour les voyelles arrondies, F2 pour les voyelles avant et F1 pour les voyelles ouvertes et centrales [MELLA, 1994]. De plus, la comparaison de phonèmes situés dans le même contexte phonémique, syntaxique et sémantique permet de minimiser l'influence de la coarticulation [MELLA, 1994 ; INGRAM ET AL., 1996].

PTACEK observe qu'avec l'âge la valeur maximale du premier formant s'abaisse et SUZUKI remarque que la valeur moyenne de F3 et F4 diminue légèrement ; SUZUKI constate aussi que le vieillissement n'amène aucune modification extrême [PTACEK ET AL., 1966 ; SUZUKI ET AL., 1994]. Par contre l'émotion affecte la position des formants, particulièrement dans la première syllabe des mots [KRAUSE, 1976].

4.4.2.1.2. Trajectoire des formants

La trajectoire des formants englobe des aspects statiques et dynamiques de la parole. Elle contient des informations sur des caractéristiques dépendantes du locuteur, comme la taille du tractus vocal et la forme des résonateurs, la stratégie d'articulation et les effets de la coarticulation et de la diphtongaison, le dialecte et l'accent ou la manière de parler [SAMBUR, 1975 ; INGRAM, 1995]. De plus, la trajectoire des deux ou trois premiers formants est relativement robuste aux différents bruits [JANKOWSKI ET AL., 1994].

4.4.2.1.3. Mesure d'énergie

Les habitudes et les pratiques d'articulation des consonnes et des voyelles d'un locuteur peuvent être évaluées par la distribution de l'énergie de certains segments dans le domaine spectral [HIRSON ET DUCKWORTH, 1993]. Dans l'ensemble des consonnes par exemple, la variabilité interlocuteur du /s/ semble supérieure à sa variabilité intralocuteur [FRENCH, 1994].

4.4.2.2. Caractéristiques segmentales temporelles

4.4.2.2.1. L'indice de régularité mélodique ou *jitter*

Le *jitter* est la mesure de la variation cycle à cycle de la période vibratoire du larynx. Cet indice s'exprime en pourcentage de la fréquence fondamentale. Plusieurs modes de calcul ont été

proposés, mais le quotient de perturbation de la fréquence fondamentale, ou *Pitch Perturbation Quotient* (PPQ), a été retenu par la plupart des auteurs [KOIKE, 1973 ; DAVIS, 1976].

$$PPQ = \frac{100}{N-1} \frac{\sum_{i=2}^{i=N} |F_{0i} - F_{0i-1}|}{F_{0\text{moyenne}}} \quad (4.1)$$

Cette caractéristique est utilisée en phonologie clinique pour quantifier la sensation subjective de raucité des voix et diagnostiquer les pathologies qui la provoquent. La plupart des méthodes permettent la mesure du *jitter* dans des conditions d'enregistrement de haute qualité à partir de voyelles isolées et soutenues, mais ne sont pas applicables aux conditions rencontrées en sciences forensiques. Un algorithme permettant l'extraction de cette caractéristique dans des enregistrements de parole spontanée et dans des conditions dégradées a été proposé par WAGNER ; les résultats obtenus n'aboutissent cependant qu'à une discrimination entre locuteurs sains et pathologiques [WAGNER, 1995].

4.4.2.2. Durées des segments

Les durées des segments peuvent être trompeuses lorsqu'elles sont isolées du contexte d'analyse, mais elles peuvent contribuer à la détermination de l'analyste, particulièrement lorsque des comportements pathologiques ou inhabituels sont mis en évidence dans les échantillons [FRENCH, 1994].

4.4.2.3. Caractéristiques suprasegmentales fréquentielles

4.4.2.3.1. Mesure de la hauteur de la fréquence fondamentale moyenne

La fréquence fondamentale n'existe, *stricto sensu*, que pour un son voisé d'origine laryngée ; elle correspond alors à la fréquence de vibration pseudo-périodique des cordes vocales. Dans le registre de poitrine ou modal, utilisé normalement pour la phonation, le locuteur utilise une fréquence fondamentale usuelle. Il la conserve dans une plage de fréquences donnée, la tessiture, inférieure aux extrêmes de l'étendue de la voix, qui peuvent être atteints avec le registre de fausset. La fréquence fondamentale peut être mesurée pour des échantillons de voix parlée, mais elle n'est pas appropriée au chant, car la hauteur de la voix a été prédéfinie par le compositeur [ORMEZZANO ET ROCH, 1991].

La mesure de la fréquence fondamentale permet de déterminer la distribution des fréquences à l'intérieur de la tessiture et d'en extraire la fréquence fondamentale modale, qui correspond à la fréquence la plus utilisée par le locuteur et la fréquence fondamentale moyenne ou *pitch*, qui est la moyenne de toutes les mesures. Ces deux valeurs sont en général très proches, mais peuvent différer en présence de pathologies [ORMEZZANO ET ROCH, 1991].

$$F_{0\text{moyenne}} = \frac{\sum_{i=1}^{i=N} F_{0i}}{N} \quad (4.2)$$

F_0 est hautement significative pour la reconnaissance de locuteurs des deux sexes, pour les auditeurs comme pour les machines [LARIVIERE, 1975 ; SAMBUR, 1975]. Les données statistiques concernant la mesure de F_0 suggèrent l'existence d'une fréquence fondamentale spécifique au locuteur, dont la distribution se stabilise lorsque la durée des échantillons croît [HORII, 1975]. 10 s de parole spontanée permettent déjà d'obtenir une distribution grossière, mais des durées de 30 s à une minute sont nécessaires pour obtenir une distribution fine et stable [STEFFEN-BATOG *ET AL.*, 1970 ; HILLER *ET AL.*, 1984].

La population féminine peut être clairement distinguée de la population masculine par la fréquence fondamentale [ATKINSON, 1976 ; COLEMAN, 1976] et la distribution de la fréquence fondamentale moyenne de la parole spontanée des locutrices (F_0 moyenne = 240 Hz, $\sigma = 40$) est approximativement le double de celle des locuteurs (F_0 moyenne = 125 Hz, $\sigma = 20.5$) (Figure IV.1.) [SAITO *ET AL.*, 1958 *IN* : FURUI, 1989].

Comme la limite inférieure de la bande passante du système téléphonique, située à 300 Hz, est supérieure à la fréquence fondamentale de la plupart des locuteurs, celle-ci est calculée à partir de ses harmoniques H_1 à H_n présentes dans le signal [KELLER, 1994].

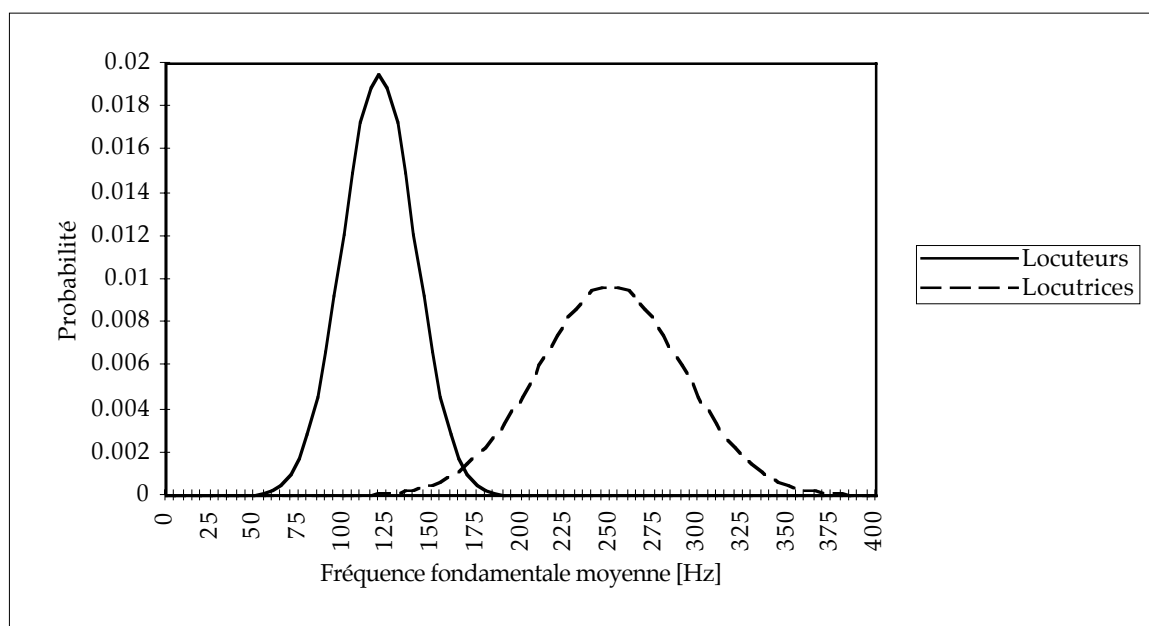


Figure IV.1. Approximation de la distribution de la fréquence fondamentale moyenne de la parole spontanée des locutrices et des locuteurs [SAITO *ET AL.*, 1958 *IN* : FURUI, 1989]

4.4.2.3.2. L'intonation

L'intonation, appelée aussi mélodie de la parole ou contour de F_0 , correspond, au niveau acoustique, à la variation temporelle à long terme de la fréquence fondamentale. L'essentiel de la recherche phonétique s'est concentré sur l'information linguistique contenue dans l'intonation et le fait que cette fonction linguistique soit de première importance ne laisse que peu de place à des caractéristiques dépendantes du locuteur. ATKINSON montre que la variabilité intralocuteur et

interlocuteur du contour de F_0 est comparable lorsqu'un seul et même type d'énoncé est utilisé [ATKINSON, 1976].

La contribution du contour de F_0 à l'identification de locuteurs familiers a cependant été mise en évidence par ABBERTON ET FOURCIN, et plus systématiquement par VAN DOMMELEN [ABBERTON ET FOURCIN, 1978 ; VAN DOMMELEN, 1987]. Ces deux études montrent que les locuteurs dont le contour de F_0 est semblable sont plus facilement confondus par des auditeurs que lorsqu'il est différent. L'importance du contour de F_0 est cependant secondaire, la plus grande partie de la tâche d'identification ayant déjà été effectuée sur la base de la hauteur de F_0 au moment où la durée de l'énoncé est suffisante pour que l'intonation soit prise en compte [VAN DOMMELEN, 1990].

D'autres investigations sur l'un des paramètres constituant le contour de F_0 , la tessiture, montrent qu'elle influence la perception que les auditeurs ont de l'état psychologique du locuteur [BROWN, 1974 ; LADD ET AL., 1985]. Cette information est donc probablement aussi codée, au même titre que l'information linguistique, ce qui limite la transmission d'information dépendante du locuteur [VAN DOMMELEN, 1990].

4.4.2.4. Caractéristique suprasegmentale temporelle

4.4.2.4.1. Le rythme

Une conversation se compose de parole et de pauses. Le temps de parole de chacun des interlocuteurs est d'environ un tiers, mais il dépend de leur vitesse d'élocution [FURUI, 1989].

Dans la vitesse d'élocution, la transmission d'informations spécifiques au locuteur est limitée, car l'organisation temporelle de la parole découle des durées spécifiques de chaque segment et des contraintes imposées par l'information linguistique, véhiculée dans l'énoncé. La plupart des études ont cependant été menées sur ce seul aspect du rythme de la parole.

Des variations de plus ou moins 10% de la vitesse globale d'élocution ont une influence significative sur le taux d'identification subjectif par des auditeurs [BROWN, 1981] et des altérations plus drastiques, de plus ou moins un tiers de la vitesse, font décroître les taux de reconnaissance de 10 à 17% [VAN LANCKER ET AL. 1985B]. Ces altérations ont aussi un effet sur la perception de l'état psychologique du locuteur par les auditeurs [BROWN ET AL., 1974]. Ces résultats indiquent que la structure temporelle est l'un des facteurs qui contribuent à la reconnaissance de locuteurs.

4.4.2.5. Pathologies

La constatation de troubles du comportement vocal peut permettre de détecter, dans une certaine mesure, la présence de dysfonctionnements organiques ou moteurs. Les atteintes pathologiques ou accidentelles des organes de la phonation, comme la présence de kystes sur les cordes vocales ou une blessure des cordes vocales causée par intubation, peuvent affecter la voix en abaissant la fréquence fondamentale ou lui donner un timbre rauque, grinçant ou discordant.

Par contre, des lésions du système nerveux central ou périphérique affectent plutôt la parole, en provoquant des difficultés de l'élocution [ORMEZANO ET ROCH, 1991 ; BRAUN, 1995]. La caractérisation de ces troubles comportementaux devrait être effectuée avec l'aide d'une personne de l'art, phoniatre, orthophoniste ou neuropsychologue.

4.4.2.6. Limites de l'approche phonétique acoustique

L'analyse phonétique acoustique peut apporter une vue quantitative, plus détaillée et plus claire de l'information contenue dans le signal de parole. Les détails des dimensions acoustiques qui sous-tendent les impressions perceptives, comme la fréquence fondamentale, les fréquences formantiques et les durées, peuvent être observés analytiquement avec une précision que les limites du système perceptif humain ne permettent pas d'atteindre.

Cependant, les caractéristiques dépendantes du locuteur et celles qui procèdent des autres fonctions du langage évoluent dans les mêmes dimensions et leur variabilité résulte d'un grand nombre d'influences. Dès lors, l'aptitude à mesurer une différence acoustique isolée n'implique pas forcément la capacité d'évaluer sa signification du point de vue de l'identification et une interprétation avertie s'avère nécessaire [NOLAN, 1991].

4.4.3. Limites des approches auditive et phonétique acoustique

L'hypothèse de l'unicité de la voix humaine réside dans la possibilité de caractériser chaque individu dans un domaine de variation unique d'un espace multidimensionnel, lorsque l'on considère un nombre de dimensions suffisant. Pourtant, cette hypothèse ne peut pas être considérée comme un fait, ni en général, ni dans le cadre de la tâche d'identification de locuteurs en sciences forensiques [NOLAN, 1991].

Les phonéticiens sont capables de livrer une analyse structurée d'une grande qualité, mais même leur grand savoir-faire n'assure pas de garantie [HOLLIEN, 1990]. De même, FRENCH reconnaît que personne ne peut établir l'identité d'un locuteur avec une certitude scientifique absolue : « Malgré l'amélioration constante de la technologie utilisée par les phonéticiens dans leurs analyses forensiques, les conclusions auxquelles ils parviennent demeurent du niveau de l'opinion et devraient être utilisées de façon corroborative » [FRENCH, 1994].

V. APPROCHE SPECTROGRAPHIQUE

5.1. Le spectrographe sonore

5.1.1. La technologie

Suite aux travaux de STEINBERG, le spectrographe sonore a été mis au point chez *Bell Telephone Laboratories*[®] en 1941, en tant qu'instrument d'analyse fondamentale de la voix, aussi bien dévolu aux applications phonétiques, à l'usage des sourds, à l'apprentissage des langues qu'à l'amélioration de la qualité des transmissions téléphoniques [STEINBERG, 1934 ; POTTER ET AL., 1947 ; ALEXANDERSON, 1997]. Cet instrument permet de représenter les variations temporelles du spectre à court terme d'une onde de parole sous une forme graphique, appelée spectrogramme vocal, vocogramme [KERSTA, 1973] ou sonagramme [NOLAN, 1983]. Le prototype de l'instrument analogique a été décliné en deux formes : le *Direct Translator* destiné aux sourds, permettant une visualisation directe du résultat sur un écran cathodique phosphoré, et la version commune du spectrographe sonore offrant le résultat sous forme imprimée.

Sur le spectrogramme, le temps occupe la dimension horizontale, les fréquences la dimension verticale et la densité du trait indique l'intensité [BOLT ET AL., 1970]. Cette représentation permet la mise en évidence de plusieurs informations contenues dans le signal de parole comme la largeur de bande et la pente des formants des voyelles, leurs fréquences centrales, la durée des événements acoustiques, les formes caractéristiques des consonnes fricatives et l'énergie entre les formants [CORSI, 1982].

Dans les années soixante, une version commerciale du spectrographe sonore, proposée par *Kay Elemetrics Corporation*[®], a été largement utilisée dans la recherche phonétique acoustique, tandis que *Voiceprint Laboratories Corporation*[®] proposait une version dotée d'un système de lecture continue des enregistrements, destinée au domaine de l'identification [PRESTI, 1966]. Dès le début des années 1980, la puissance de calcul offerte par les processeurs et la disponibilité de moniteurs vidéo de haute définition ont rendu possible la réalisation de stations de travail informatiques dotées d'une capacité de visualisation spectrographique complète³⁵ ; elles sont maintenant financièrement accessibles et présentes dans tous les laboratoires [GRUBER ET POZA, 1995].

5.1.2. L'application à la reconnaissance de locuteurs

La participation des États-Unis à la deuxième guerre mondiale a donné naissance à un projet d'application militaire du spectrographe sonore : l'identification de navires ennemis par l'intermédiaire de la voix de leurs opérateurs radio. A cause de l'intérêt militaire, la publication

³⁵ *infra* : 6.2.2.2.3. Spectrogrammes numériques par transformée de Fourier rapide

d'informations concernant la nature et l'avancée de ce projet a été ajournée jusqu'à la fin de la guerre [POTTER, 1946]. En 1946, les inventeurs du spectrographe sonore précisent que si les spectrogrammes vocaux comportent certaines caractéristiques dépendantes du locuteur, les ressemblances sont beaucoup plus grandes que les différences lorsque deux locuteurs différents prononcent un même énoncé ; ils mentionnent en outre que les recherches concernant l'identification de locuteurs ne sont pas terminées [KOPP ET GREEN, 1946]. Seule une brève publication suggère l'utilisation des spectrogrammes vocaux comme méthode d'identification d'un point de vue légal [STEINBERG ET FRENCH, 1946].

Les recherches continuent aux *Bell Telephone Laboratories*[®] durant seize ans sous la direction de KERSTA, mais l'absence de toute publication entre 1946 et 1962 laisse à penser que le projet était classifié et financé par l'armée [ALEXANDERSON, 1997 ; SMRKOVSKI, 1997].

Dans le contexte de la guerre froide, la connaissance des recherches nord-américaines conduit les autorités de l'Union Soviétique à lancer un programme de recherche sans précédent en 1949 à Mavrino, près de Moscou, dans une prison spéciale du régime réservée aux prisonniers politiques et idéologiques. La plupart des détenus sont des ingénieurs et des techniciens pour lesquels la prison représente le premier cercle de disgrâce et le camp de déportation le dernier [SOLZENICYN, 1968]. Pour avoir vécu à Mavrino et participé au programme de recherche sur l'identification de locuteurs, le prix Nobel de littérature 1970, Aleksandr Isaevic SOLZENICYN, alors professeur de physique à Riazan, rend compte de ce programme avec de très grands détails dans son roman « Le Premier Cercle », écrit de 1955 à 1958, confisqué en Union Soviétique et édité en France où il a reçu le Prix du meilleur livre étranger en 1968.

5.2. L'application forensique

5.2.1. La méthode de KERSTA

5.2.1.1. Bases théoriques, méthodologie et résultats

Lorsque KERSTA propose l'utilisation du spectrographe sonore dans le domaine forensique à *Bell Telephone Laboratories*[®], l'entreprise répond qu'elle est une compagnie de téléphone et refuse le développement de cette application. Par contre, elle permet à KERSTA de prendre une retraite anticipée et d'emporter avec lui la technologie du spectrographe sonore [ALEXANDERSON, 1997]. En 1962, KERSTA publie la méthode d'identification de locuteurs par comparaison visuelle de spectrogrammes vocaux dans deux articles intitulés « *Voiceprint Identification* » publiés dans les revues *Nature* et *Journal of Acoustical Society of America* [KERSTA, 1962A ; KERSTA, 1962B]. A la fin de l'année, il donne aussi une conférence appelée « *Voiceprint Identification Infallibility* » devant *l'Acoustical Society of America*.

Pour nommer les spectrogrammes vocaux, KERSTA reprend la dénomination d'« empreinte vocale » ou « *voiceprint* », terminologie utilisée de façon interne dans les laboratoires *Bell*, par analogie aux empreintes digitales [GRAY ET KOPP, 1944 IN : TOSI ET AL., 1972B]. Fort de cette analogie fallacieuse, il propose l'utilisation des spectrogrammes vocaux en sciences forensiques et leur prête

les caractéristiques d'individualité des empreintes digitales, par le fait que, d'une part, l'anatomie des cavités vocales et les points d'articulation varient chez chaque être humain et que, d'autre part, leur contrôle par le système nerveux diffère d'un individu à l'autre [LADEFOGED ET VANDERSLICE, 1967 ; BOLT ET AL., 1970]. Il reprend à son compte l'hypothèse théorique de la « parole invariante » développée par les phonéticiens FANT et LADEFOGED, qui présume que la variabilité intralocuteur des caractéristiques spectrales de la parole est inférieure à la variabilité interlocuteur et KERSTA tente de démontrer cette hypothèse par une étude sur une population de 12 locuteurs [FANT, 1960 ; LADEFOGED, 1962 ; KERSTA, 1962A].

Par analogie encore, KERSTA attribue aux « empreintes vocales » la probabilité de coïncidence fortuite estimée pour les empreintes digitales et la possibilité d'identification formelle qui en découle. Il effectue des analyses uniquement visuelles, avec dix mots cibles très courants, *a, and, I, is, it, me, on, the, to* et *you*, enregistrés de manière isolée. Pour des examinateurs entraînés, le taux d'erreur est évalué à environ 1% avec l'utilisation d'un seul de ces mots, erreur que KERSTA compare à celle existant lors de l'identification par empreinte digitale avec un seul doigt, ce qui est incorrect [CHAMPOD, 1996], et conclut qu'avec l'utilisation de plusieurs mots ce taux d'erreur diminue [KERSTA, 1962A]. Une autre étude montre que les performances de l'examineur augmentent lorsqu'il peut comparer plusieurs mots cibles simultanément ; elle indique aussi que les performances de la méthode sont, en moyenne, comparables pour la reconnaissance des hommes et des femmes et que les performances ne sont pas affectées par le déguisement de la voix ou par le passage de la voix à travers le canal téléphonique [ANONYME, 1965 IN : HECKER, 1971].

5.2.1.2. Témoignages en cour

En 1966, KERSTA quitte *Bell Telephone Laboratories*[®] et fonde l'entreprise *Voiceprint Laboratories Corporation*[®] à Sommerville, NJ, qui vend l'équipement spectrographique et offre une formation en identification de locuteurs. La même année, la première expertise de KERSTA, basée sur la technique de la comparaison visuelle des « empreintes vocales », est admise par une cour de justice, dans l'affaire *People v Straehle*³⁶. Cette cour interprète le silence de la communauté scientifique au sujet des travaux de KERSTA comme une acceptation générale tacite et conclut que cette méthode nouvelle est admissible puisqu'elle satisfait au standard de *Frye*³⁷.

Une formulation de ce standard, souvent citée³⁸, précise que la cour doit distinguer « l'étape expérimentale de l'étape de démonstration d'une découverte ou d'un principe scientifique et qu'elle ne devrait admettre un tel témoignage que lorsque le principe, duquel la déduction est tirée, est suffisamment établi pour avoir gagné une acceptation générale de la communauté scientifique pertinente » [LOEVINGER, 1995].

En 1967 dans l'affaire *United States v Wright*³⁹, la cour Militaire d'Appel des États-Unis condamne un membre de l'Air Force pour des appels téléphoniques anonymes menaçants et

³⁶ [People v Straehle, No 9323/64 (Sup. Ct Westchester County), noté dans 12 New York L F 501 (1966)]

³⁷ *supra* : 3.3. Exigences légales en matière de preuve scientifique

³⁸ [State v Valdez, (1962) 91 Ariz 274, 371 P894]

³⁹ [United States v Wright, 17 CMA 183, 37 MR 447]

obscènes, sur la base de l'identification effectuée par KERSTA. Elle est cependant la seule cour d'appel à avoir admis la validité de la méthode, uniquement sur la base des expériences de KERSTA [GOCKE ET OLENIEWSKI, 1973]. La cour a cependant mal identifié le principe scientifique duquel la déduction est tirée. Il ne s'agit ni du spectrographe ni de son produit, le spectrogramme, puisque aucun des deux n'est doté d'une capacité de reconnaissance de forme, mais il s'agit de l'hypothèse que la voix est unique, que cette unicité est reproduite sur le spectrogramme et qu'elle peut être détectée par l'œil humain. Malheureusement la plupart des cours ont statué sur l'acceptation par la communauté scientifique de la fiabilité et de la reproductibilité du spectrographe [THOMAS, 1981].

Les deux principaux cas de cette période initiale sont jugés en 1968, *State v Cary*⁴⁰ et *People v King*⁴¹, avec comme base de la décision juridique le précédent constitué par l'affaire *United States v Wright*⁴² [BOLT ET AL., 1970]. Cependant les cours d'appel du New Jersey et de Californie rejettent la méthode, statuant que le processus d'identification par « empreintes vocales » n'a pour l'heure ni atteint une acceptation scientifique suffisante, ni prouvé sa fiabilité pour être admis comme preuve d'identification dans des cas où la vie ou la liberté d'une personne est en jeu [KENNEDY, 1968 IN : HECKER, 1971]. Cette décision s'appuie largement sur les positions des représentants de la communauté scientifique pertinente, au sens du standard de *Frye* [LADEFOGED ET VANDERSLICE, 1967 ; MCDADE, 1968].

Suite à ces deux affaires, KERSTA, qui avait témoigné dans huit affaires, dont sept fois pour l'accusation, interrompt son activité d'expert entre 1968 et 1970.

5.2.1.3. Prise de position de la communauté scientifique et juridique sur l'étude de KERSTA

Dans le *New York Times* du 12 avril 1966, BORDERS fait remarquer que la faillibilité de l'observateur est un problème crucial pour l'utilisation légale de la méthode spectrographique et auditive [BORDERS, 1966]. Dès 1967, les résultats des expériences menées par KERSTA sont contestés et infirmés par YOUNG ET CAMPBELL, qui montrent que pour un échantillonnage de cinq locuteurs, le taux d'identification correcte est de 78,4% en cas d'utilisation de mots isolés, mais seulement de 37,3% en cas d'utilisation de mots extraits d'énoncés de parole spontanée [YOUNG ET CAMPBELL, 1967]. STEVENS et ses collègues montrent que les performances sont proportionnelles à la durée des énoncés et que les performances de la méthode de reconnaissance auditive surpassent systématiquement celles obtenues par la méthode de comparaison visuelle de spectrogrammes [STEVENS ET AL., 1968]. Pour l'identification de huit locuteurs, elles sont de 88% pour la méthode auditive contre 68% pour la méthode visuelle à partir d'une seule syllabe, de 90% contre 75% à partir de mots isolés et de 91% contre 83% à partir d'une phrase entière. Dans le même temps HECKER met en évidence les modifications significatives des spectrogrammes provenant d'une voix

⁴⁰ [State v Cary, (1967) 49 N.J. 343]

⁴¹ [People v King, (1968, 2nd Dist) 266 Cal App 2d 437, 72 Cal Rptr 478]

⁴² [United States v Wright, 17 CMA 183, 37 MR 447]

générée dans des conditions de stress, par rapport à la même voix, générée dans des conditions normales [HECKER *ET AL.*, 1968].

YOUNG ET CAMPBELL, ainsi que STEVENS, montrent aussi que certains locuteurs sont considérablement plus difficiles à reconnaître par leurs spectrogrammes que d'autres [YOUNG ET CAMPBELL, 1967 ; STEVENS *ET AL.*, 1968]. A cause de la grande variabilité de cette capacité d'identification, le groupe de locuteurs de test devrait être aussi grand que possible et homogène d'un point de vue de la perception auditive. HECKER mentionne qu'il n'existe que peu d'informations concernant les corrélations entre les ressemblances perceptuelles auditives et les ressemblances spectrographiques visuelles [HECKER, 1971]. De plus, ce critère d'homogénéité auditive lui semble inapproprié et il ne relève pas que la proximité auditive de deux voix est à la base de toute décision préalable de procéder à une expertise forensique. ROTHMAN a d'ailleurs montré une réduction des performances d'identification par comparaison visuelle de spectrogrammes lorsque la proximité auditive de deux échantillons est grande [ROTHMAN, 1979].

En Allemagne, ENDRESS *ET AL.* testent la variation des spectrogrammes en fonction de l'âge, du déguisement et de l'imitation de la voix, dans le but de vérifier les hypothèses développées par KERSTA [ENDRESS *ET AL.*, 1971]. Les tests réalisés avec six locuteurs d'âge variant entre 29 et 43 ans offrent des résultats sans équivoque : ni la structure des formants des voyelles, ni la fréquence fondamentale ne sont indépendants de l'âge. La possibilité de modifier la structure des formants des voyelles et la fréquence fondamentale est considérable grâce à un déguisement délibéré de la voix. En cas d'imitation, les caractéristiques imitées permettent d'associer de manière auditive la voix de l'imitateur à celle de la personne imitée, mais ces caractéristiques sont difficiles à définir et à localiser sur les spectrogrammes vocaux.

Seul TOSI, professeur au Département d'Audiologie et des Sciences de la Parole de l'Université d'État du Michigan, considère comme prometteuses les études de KERSTA, malgré le besoin d'études complémentaires indépendantes spécifiques [TOSI, 1967 ; TOSI, 1968]. Malheureusement, son enthousiasme est déjà démenti par la première étude publiée par l'École de Justice Criminelle de l'Université d'État du Michigan. HENNESSY ET ROMIG réexaminent la méthode de comparaison visuelle de spectrogrammes d'un point de vue théorique ; ils reproduisent aussi les expériences de KERSTA, mais les taux d'identification qu'ils mettent en évidence ne sont que de 70% en moyenne. Ils en concluent que la validation de cette méthode n'a pas été faite et que la controverse qui l'entoure ne pourra être levée que par la démonstration de l'hypothèse d'individualité des spectrogrammes [HENNESSY ET ROMIG, 1971A ; HENNESSY ET ROMIG, 1971B].

5.2.1.4. Rapport du *Technical Committee on Speech Communication of the Acoustical Society of America* (BOLT I)

Devant l'incapacité de KERSTA à publier des rapports techniques détaillés et face à l'abondance des publications motivées par cette controverse [CEDARBAUMS, 1969 *IN* : CUTLER *ET AL.*, 1972 ; KAMINE, 1969 *IN* : CUTLER *ET AL.*, 1972], le *Technical Committee on Speech Communication of the Acoustical Society of America* charge six chercheurs dans le domaine de la parole, BOLT, COOPER, DAVID, DENES, PICKETT et STEVENS, de considérer les problèmes suivants :

1. « Lorsque deux spectrogrammes vocaux se ressemblent, cela signifie-t-il plutôt 'même locuteur' ou 'même mot prononcé' ? »
2. « Les ressemblances non pertinentes sont-elles de nature à induire un jury en erreur lors de l'évaluation des témoignages de deux experts opposés ? »
3. « Dans quelle mesure les spectrogrammes sont-ils dépendants du locuteur ? »
4. « Quelle est la variation temporelle des spectrogrammes ? »
5. « Sont-ils sensibles au déguisement de la voix ou susceptibles d'être falsifiés ? »

Dans leur réponse, connue sous le nom de « BOLT I », les auteurs mettent à jour les lacunes théoriques et expérimentales de la démonstration de l'hypothèse de KERSTA [BOLT ET AL., 1970] :

1. « La parole charrie plusieurs messages très interdépendants, simultanément entremêlés de façon complexe. Les caractéristiques dépendantes du locuteur sont difficiles à dégager car elles ne sont pas connues. Cependant, dans une certaine mesure, l'humain peut réaliser cette tâche, de façon auditive ou par observation de spectrogrammes. Le signal acoustique de la voix peut être analysé en fréquence, en énergie et en temps et visualisé sous forme de spectrogramme, mais aucune représentation ne visualise directement des traits individuels de la voix à cause de leur mélange. La décision d'identification demeure subjective ».
2. « Les similarités et les différences existant sur les spectrogrammes sont ambiguës et peuvent être mal interprétées. L'analogie entre empreintes digitales et vocales est fallacieuse et l'interprétation liée à ces deux indices est très différente. Le dessin des empreintes digitales et palmaires sont inhérents à l'anatomie et immuables. Seule la destruction du derme peut affecter leur structure. Les détails de ces dessins, les minuties, sont permanents et ne sont pas affectés par la croissance et les habitudes ; seuls la taille et le grain de la peau évoluent. Les dessins dépendent du résultat d'un transfert direct de la peau du doigt sur une surface qu'il a touchée. Les sons produits par la voix dépendent en premier lieu d'attitudes apprises pour produire le code du langage et seulement partiellement de la structure anatomique. Les spectrogrammes vocaux résultant d'une analyse de ces sons sont modifiés par les mouvements articulatoires exigés pour réaliser le code du langage et sont seulement indirectement corrélés à l'anatomie du locuteur. Les détails de ces dessins sont affectés par la croissance et les habitudes, par les connaissances et l'état de santé du locuteur. De plus le canal de transmission, du locuteur au spectrographe, est vulnérable aux distorsions acoustiques et électriques ».
3. « Les performances de la méthode dépendent de la tâche, des circonstances et de l'examineur ».
4. « Les études menées dans ce domaine ne sont pas suffisantes pour aboutir à une évaluation objective ».
5. « Les méthodes, les procédures et les tests de validité n'ont pas été publiés par les auteurs. ».

5.2.1.5. Prise de position du *Federal Bureau of Investigation* (FBI)

Dans une lettre à l'éditeur du *Journal of Criminal Law, Criminology and Police Science* du 21 décembre 1971, le directeur du FBI, Edgar J. Hoover, révèle la position prudente de l'agence vis-à-vis de cette technique :

« Nous estimons que la comparaison des empreintes vocales est utile à des fins d'investigation, mais pour l'instant son authenticité n'a pas été suffisamment établie ni démontrée pour servir de base fiable au témoignage d'un expert et à l'identification ».

5.2.2. Tentative de validation de la méthode de KERSTA : l'étude de TOSI

En 1968, cette situation controversée contraint le *Law Enforcement Administration Assistance of the United States Department of Justice* (LEAA) à commander une revue bibliographique complète et exégétique du domaine de la reconnaissance de locuteurs au *Sensory Sciences Research Center of the Stanford Research Institute* (SRI) de Menlo Park, Californie, sous l'égide de l'*American Speech and Hearing Association* (ASHA) [HECKER, 1971].

Sa parution, en janvier 1971, décide le LEAA à allouer un fonds de 300'000 dollars aux recherches sur la reconnaissance de locuteurs. Il est accordé pour moitié au SRI, pour l'étude de la reconnaissance automatique de locuteurs⁴³, et pour la seconde moitié au *Department of Michigan State Police* afin de procéder à la vérification des hypothèses de KERSTA. La validation de la méthode de KERSTA est confiée au professeur TOSI. Ce projet inclut une étude de l'identification visuelle des spectrogrammes vocaux dans des conditions forensiques réelles, dont la responsabilité est confiée au Det. Sgt. Ernest NASH, de la police d'Etat du Michigan, et technicien en identification de voix formé par KERSTA en 1966 [TOSI ET AL. 1972A ; GRUBER ET POZA, 1995]. Cette étude est la seule réalisée à grande échelle dans le domaine de l'identification visuelle de spectrogrammes vocaux ; toutes les autres ont été effectuées à petite échelle et leurs méthodologies sont si différentes que les résultats sont très difficiles à comparer [BOLT ET AL., 1979].

5.2.2.1. Évaluation principale en laboratoire

L'étude de TOSI analyse l'influence de sept variables sur les performances d'identification par comparaison visuelle de spectrogrammes vocaux dans un ensemble de 250 locuteurs.

5.2.2.1.1. Variables analysées

1. Le nombre de mots cibles utilisés pour l'identification : les mots utilisés sont *it, is, on, you, and, the, I, to* et *me*, très courants en anglais. Les tests ont été effectués avec (a) neuf mots cibles et (b) six mots cibles.
2. Le nombre d'énoncés du même mot cible produit par chaque locuteur : les tests ont été effectués avec (a) une occurrence, (b) deux occurrences et (c) trois occurrences.

⁴³ *infra* : 6.4.1. Semi-Automatic Speaker Identification System (SASIS) - USA (1971 - 1975)

3. Les conditions d'enregistrement et de transmission des mots cibles : les enregistrements ont été effectués (a) dans des conditions de haute fidélité, directement avec un enregistreur et un microphone dans un environnement calme, (b) dans des conditions téléphoniques et un environnement calme à l'aide d'un combiné téléphonique muni d'un couplage acoustique et (c) dans des conditions téléphoniques et un environnement bruyant, 50 dB de bruit blanc mesurés au niveau de la tête du locuteur.
4. Le contexte des mots cibles utilisés pour l'identification : (a) énoncés de manière isolée, (b) énoncés dans un contexte fixe et (c) énoncés dans des contextes libres pour lesquels différentes phrases ont été comparées.
5. La taille de l'ensemble de test : comprenant soit (a) dix locuteurs, (b) vingt locuteurs ou (c) quarante locuteurs.
6. La variation intralocuteur temporelle : Les spectrogrammes de test et de comparaison sont issus soit (a) de la même session d'enregistrement, soit (b) de deux sessions distantes d'un mois au minimum. Contrairement au déroulement chronologique forensique, les spectrogrammes de test ont été recueillis après ceux de comparaison.
7. La tâche confiée à l'examineur est soit une tâche d'identification en ensemble fini, soit une tâche d'identification en ensemble infini. Lors de la tâche d'identification en ensemble fini (a), un échantillon du locuteur testé est présent dans l'ensemble de référence des locuteurs, alors que pour la tâche d'identification en ensemble infini, l'échantillon du locuteur testé est présent dans la moitié des cas (b) ; dans l'autre moitié il est absent (c). Évidemment, l'examineur ne possède que l'information ensemble fini ou ensemble infini [TOSI ET AL., 1972A].

5.2.2.1.2. Échantillonnage

Deux cent cinquante locuteurs masculins ont été sélectionnés aléatoirement dans la population des étudiants de l'Université d'État du Michigan. Tous sont natifs des États-Unis, de langue maternelle anglaise américaine, sans défaut de parole ni différence dialectale marquée. L'âge des locuteurs s'étend de 17 à 27 ans, avec une moyenne à 19,8 ans et un écart type de 2,1 ans [TOSI ET AL., 1972A].

Quant aux examinateurs, ils ont été recrutés par voie d'annonce et sélectionnés après une brève explication de la méthode spectrographique et deux tests éliminatoires, portant sur l'identification visuelle d'un locuteur dans un ensemble de trois, puis de onze personnes. Avant de commencer l'expérimentation, les 29 examinateurs choisis suivent un mois de cours de phonétique et de lecture de spectrogrammes ainsi qu'un entraînement dans la tâche d'identification en ensemble fermé.

Les examinateurs ont été placés dans trois groupes : le premier (I) composé de femmes de 17 à 60 ans, de différentes formations, le suivant (II) composé d'étudiants non diplômés de différentes facultés de l'Université d'État du Michigan et le dernier (III) composé uniquement d'étudiants du département de Justice Criminelle de cette université. En plus, trois sous-groupes d'un, de deux ou de trois examinateurs ont été formés. Le cycle d'identification à partir de neuf mots cibles (1a) a été

réalisé par les 29 examinateurs, alors que le second cycle à partir de six mots cibles (1b) n'a été réalisé que par les 15 examinateurs les plus motivés et les plus doués [TOSI ET AL., 1972A].

5.2.2.1.3. Résultats

Variable étudiée		Contexte	1er cycle (1a)			2ème cycle (1b)		
			Ens. fini	Ensemble infini		Ens. fini	Ensemble infini	
			Er. type I	Er. type II	Er. type I	Er. type I	Er. type II	Er. type I
2a	Un énoncé	inconnu	8,71 %			10,29 %		
2b	Deux énoncés	inconnu	9,04 %			8,38 %		
2c	Trois énoncés	inconnu	7,61 %			7,51 %		
3a	Haute-fidélité	inconnu	7,58 %			8,59 %		
3b	Env. tél. calme	inconnu	8,69 %			7,80 %		
3c	Env. tél. bruité	inconnu	8,98 %			8,90 %		
5a	10 locuteurs	inconnu	6,97 %			6,17 %		
5b	20 locuteurs	inconnu	8,13 %			8,32 %		
5c	40 locuteurs	inconnu	10,42 %			11,80 %		
6a	Contemporain	m. isolés (4a)	0,51 %	0,36 %	1,23 %	0,62 %	0,52 %	1,70 %
		c. fixe (4b)	1,03 %	1,49 %	1,55 %	1,34 %	1,09 %	2,31 %
		c. libre (4c)	7,51 %	4,01 %	8,28 %	6,38 %	1,96 %	10,34 %
6b	Non contemp.	m. isolés (4a)	2,47 %	2,37 %	7,25 %	5,66 %	4,22 %	9,01 %
		c. fixe (4b)	9,67 %	4,22 %	10,13 %	9,88 %	4,27 %	12,68 %
		c. libre (4c)	11,83 %	6,43 %	11,83 %	10,39 %	4,81 %	10,29 %
7a	Connaissance	inconnu	5,52 %			5,69 %		
7b	Non connais.	inconnu		9,86 %			10,29 %	

Tableau V.1. Synthèse des résultats de l'étude de TOSI [TOSI ET AL., 1972B]

La combinaison de toutes les variables analysées offre un maximum de 972 ($2^2 \times 3^5$) conditions expérimentales. Le tableau V.1. reprend tous les résultats disponibles dans [TOSI ET AL. 1972A ; TOSI ET AL., 1972B], mais ceux-ci sont lacunaires. En effet seuls les résultats de 46 conditions expérimentales sont présentés (Tableau V.1.) et il est impossible d'identifier clairement ces conditions dans les commentaires qui leur sont joints.

5.2.2.2. Évaluation dans des conditions forensiques réelles

Si deux chapitres sont consacrés aux résultats de l'évaluation principale en laboratoire, les résultats de l'évaluation dans des conditions forensiques réelles menée par NASH sont curieusement placés après la discussion et les conclusions dans un chapitre intitulé : « *Extension of Results From Forensic Models to Real Cases* » [TOSI ET AL., 1972B]. Les deux extraits reproduits ci-dessous illustrent de manière patente que les conclusions reposent sur une méthodologie

discutable qui conduit à des inférences douteuses, voire fausses ⁴⁴ ; en tout cas aucune ne repose sur une évaluation empirique.

5.2.2.2.1. Échantillonnage et résultats

« Sur un total de 673 affaires, une identification positive a été obtenue dans 88 cas. Plus tard, la plupart des accusés ont avoué leur culpabilité ou ont été convaincus de culpabilité sur la base d'autres éléments. Les autres cas se répartissent en 172 décisions d'élimination positive, 31 cas d'identification ou élimination possible et 382 cas d'impossibilité de conclure à cause de la faible quantité ou qualité des échantillons » [TOSI ET AL., 1972B].

5.2.2.2.2. Extension des résultats

« Dans les cas forensiques, l'ensemble de référence des voix peut théoriquement contenir des millions d'échantillons... Cependant ce n'est pas le cas dans les situations pratiques actuelles de la police. L'ensemble de référence est théoriquement infini, bien sûr, mais pratiquement limité à un petit nombre de suspects. Il semble donc raisonnable de penser que l'intravariabilité et l'intervariabilité du groupe de suspects ne différera pas notablement des variabilités existantes dans le groupe très homogène des locuteurs expérimentaux utilisés dans cette étude » [TOSI ET AL., 1972B].

« Dans les cas forensiques, l'examineur professionnel peut normalement utiliser le temps nécessaire à l'obtention d'une conclusion et il est conscient de la conséquence d'une fausse décision. Il est donc raisonnable de conclure que les différences entre examinateurs expérimentaux et professionnels aident à améliorer les performances des seconds » [TOSI ET AL., 1972B].

« Dans les cas forensiques, les examinateurs professionnels peuvent choisir entre les décisions suivantes :

(a) Identification positive	(c) Identification possible	(e) Impossibilité de conclure
(b) Élimination positive	(d) Élimination possible	

Ce choix de décisions peut conférer une fiabilité extrême aux identifications et éliminations positives » [TOSI ET AL., 1972B].

5.2.2.3. Conclusion de l'étude de TOSI

5.2.2.3.1. Conclusion de l'évaluation principale en laboratoire

Après une analyse statistique de l'influence de chacune des sept variables analysées, LASHBROOK propose de répondre aux huit questions posées dans le projet initial [LASHBROOK, 1972] :

- Q1. « Les spectrogrammes des mêmes mots prononcés par un même locuteur dans différentes circonstances sont-ils suffisamment semblables pour être identifiés? »
- R1. « Le pourcentage de réponses correctes est en moyenne de 84,72 %. Ce résultat combine les performances en ensemble fermé (92,01%) et en ensemble ouvert (77,44%). »

⁴⁴ *infra* : 5.2.2.2.2. Extension des résultats

- Q2. « La méthode est-elle limitée par la durée séparant la production des énoncés de référence et des énoncés de test? »
- R2. « La capacité des examinateurs à identifier des locuteurs à partir de spectrogrammes contemporains est significativement supérieure (95,21 %) à l'identification à partir de spectrogrammes non contemporains (87,95 %) ».
- Q3. « Les spectrogrammes du même locuteur disant les mêmes mots sont-ils suffisamment différents des spectrogrammes de n'importe qui d'autre ? ».
- R3. « La comparaison de spectrogrammes contemporains dans un ensemble fermé occasionne une erreur d'identification de 3,05 %, alors que la comparaison de spectrogrammes non contemporains provoque une erreur d'identification de 7,99%. Dans un ensemble ouvert, les erreurs d'identification sont respectivement de 8,99% et de 22,57% ».
- Q4. « Le nombre d'énoncés du même mot utilisé pour l'identification de locuteurs altère-t-il la proportion d'identification correcte? Si oui, dans quelle mesure ? ».
- R4. « Aucune différence significative entre les pourcentages d'identification correcte ne peut être attribuée uniquement au nombre d'énoncés prononcés. Les pourcentages mis en évidence sont de 91,29% pour un énoncé, 90,96% pour deux énoncés et 92,49% pour trois énoncés ».
- Q5. « Le nombre d'échantillons de locuteurs différents présents dans l'ensemble de référence altère-t-il la proportion d'identification correcte ? Si oui, dans quelle mesure ? ».
- R5. « Les résultats indiquent une différence significative en termes d'identification correcte lorsque le nombre de locuteurs présents dans l'ensemble de référence augmente. Lorsque cet ensemble contient dix locuteurs, le pourcentage d'identification correcte est de 93,30%, lorsque l'ensemble contient 20 locuteurs il est de 91,87%, lorsque l'ensemble contient 40 locuteurs il est de 89,58% ».
- Q6. « Le pourcentage de réponses correctes dépend-il de la présence d'un échantillon du locuteur testé dans l'ensemble de référence ? ».
- R6. « Oui, l'absence d'un échantillon du locuteur testé dans l'ensemble de référence altère les performances ».
- Q7. « Le pourcentage de réponses correctes obtenu par des examinateurs entraînés dépend-il des conditions et du contexte d'enregistrement et de transmission des mots cibles utilisés pour l'identification ? ».
- R7. « L'analyse ne met pas en évidence de différence significative de pourcentage d'identification correcte lorsque le contexte d'enregistrement et de transmission change. Il est de 92,42% lorsque les enregistrements proviennent directement d'un enregistreur de bandes magnétiques, il est de 91,31% lorsque les enregistrements ont été effectués à travers le réseau téléphonique et dans un environnement calme et de 91,02% lorsque les enregistrements ont été effectués à travers le réseau téléphonique et dans un environnement bruyant ».
- Q8. « Une personne entraînée est-elle capable de reconnaître si des spectrogrammes proviennent de la même personne ou non ? ».
- R8. « En général, un examinateur entraîné est clairement capable de reconnaître les spectrogrammes des mêmes mots produits par le même locuteur. De plus, lorsque des

erreurs sont commises, le taux de faux négatifs (erreur de type II) est plus important que le taux de faux positifs (erreur de type I) pour un examinateur entraîné ».

5.2.2.3.2. Conclusion de l'évaluation dans des conditions forensiques réelles

TOSI conclut :

« Sur la base des résultats de la présente étude, des observations du travail effectué dans des conditions forensiques réelles, le Département de Justice devrait encourager l'entraînement d'experts en identification de voix, qui devraient être soigneusement testés et certifiés avant d'être reconnus comme experts par les Cours des États-Unis. Ce personnel qualifié continuera de fournir un service précieux même si une machine de reconnaissance de voix est développée dans le futur » [TOSI *ET AL.*, 1972B].

TURNER ajoute six remarques concernant l'utilisation de la méthode par la justice criminelle [TURNER *ET AL.*, 1972] :

1. « La technique d'identification par « empreintes vocales » est relativement nouvelle et huit ans est une durée très courte pour aboutir à son acceptation devant les tribunaux ».
- 2(a). « L'étude de TOSI *ET AL.*, soigneusement contrôlée, produit des taux d'erreur de 0,51% dans un ensemble fermé. D'autres expériences aboutissent à des taux d'erreur jusqu'à 29,1%, selon les conditions d'expérimentation ».
- (b). « Dans son '*Extension of Results From Forensic Models to Real Cases*' TOSI précise qu'étant donné les circonstances dans lesquelles l'investigation d'un cas réel est menée, un examinateur proprement entraîné peut s'attendre à atteindre un taux d'erreur de 1% ».
3. « L'étude de TOSI *ET AL.* indique que des examinateurs entraînés de seconde génération peuvent produire un taux d'erreur acceptable (1%) dans leur travail ».
4. « Les études HENNESSY ET ROMIG montrent que des examinateurs de seconde génération formés par apprentissage travaillant dans des conditions non contrôlées et n'utilisant aucun autre équipement que le spectrographe, n'aboutissent pas à des taux d'identification acceptables (70% et 59%) » [HENNESSY ET ROMIG, 1971A ; HENNESSY ET ROMIG, 1971B].
5. « Un entraînement approprié des examinateurs est la pierre angulaire d'une utilisation satisfaisante de la technique d'identification par 'empreintes vocales' ».
6. « La formation et l'entraînement des examinateurs amènent les recommandations suivantes :
 - (a) « Idéalement, l'expert en identification d'empreintes vocales devrait être détenteur d'une licence en science physique ou en science de la parole. Les laboratoires de sciences forensiques requièrent généralement cette exigence ».
 - (b) « Bien qu'il ait été démontré que des examinateurs entraînés de seconde génération peuvent être recrutés dans la population générale, des techniciens avec une expérience d'identification comparable devraient être préférés ».
 - (c) « En l'absence de licence, des cours de phonétique acoustique, de science de la parole, de linguistique, d'audiologie et d'électronique de base sont fortement recommandés avant toute utilisation de la technique d'identification par 'empreintes vocales' ».
 - (d) « Un entraînement minutieux dans la préparation des bandes magnétiques et des spectrogrammes vocaux est essentiel ».

- (e) « Un programme d'entraînement soigneusement supervisé d'identification de voix par comparaison de spectrogrammes doit être mené jusqu'à ce que le stagiaire atteigne un taux d'identification de 99% lors d'analyses en ensemble fermé ».
- (f) « A la suite de la formation, le stagiaire doit poursuivre son apprentissage en analysant des cas réels avec un superviseur expérimenté. Celui-ci indiquera le moment où il sentira que l'élève est suffisamment qualifié pour prendre des décisions de son propre chef ».

5.2.2.4. Création de l'International Association of Voice Identification (IAVI)

La confiance retrouvée grâce aux résultats de l'étude de TOSI *ET AL.*, KERSTA fonde l'*International Association of Voice Identification* (IAVI) avec comme président NASH jusqu'à sa retraite, puis SMRKOVSKY, et TOSI comme vice-président. Cette organisation a pour but l'entraînement, la qualification des examinateurs et la certification au rang d'expert des membres ayant accompli toutes les étapes de qualification proposées par l'association [THOMAS, 1981].

KERSTA ET NASH développent à nouveau les arguments d'individualité de la voix humaine, puisque produite par un tractus vocal de morphologie unique et confirmée par les résultats de l'étude de TOSI, et insistent sur la nature objective de l'analyse [KERSTA ET NASH, 1973 ; NASH, 1973 (curieusement publié dans le *Journal of the Association of the Official Analytical Chemists*)].

TOSI ET NASH complètent les conclusions de leur étude en montrant les différences entre les conditions expérimentales de l'étude et les conditions rencontrées par les examinateurs dans des cas réels, où idéalement l'examineur dispose de tout le temps nécessaire pour l'analyse, il est formé de manière adéquate et est conscient des conséquences de sa décision et où il lui est possible de rendre des décisions inconcluantes [TOSI ET NASH, 1973]. Les auteurs prétendent que le respect de ces conditions permet de diminuer le taux de fausses identifications et que la technique peut être utilisée à des fins d'identification si les standards suivants sont respectés :

« L'examen doit comporter une comparaison auditive et une partie visuelle. L'examineur doit être qualifié professionnellement dans les domaines de la phonétique et des sciences de la parole. Il doit éviter de se prononcer positivement s'il a le moindre doute, il doit pouvoir utiliser autant de temps que nécessaire pour effectuer les contrôles suffisants pour confirmer une conclusion. L'IAVI est habilitée à juger des qualifications professionnelles des examinateurs, à faire passer des tests pour ceux qui veulent obtenir la qualification d'expert et se donne comme mission d'encourager la recherche dans ce domaine et de renforcer le code d'éthique ».

5.2.2.5. Prise de position de la communauté scientifique et juridique sur l'étude de TOSI

Dès 1973, plusieurs études contredisent les résultats et les conclusions de l'étude de TOSI. Bien que l'échantillonnage soit plus restreint, les résultats de l'étude de HAZEN montrent que le cumul des deux types d'erreur (I et II) se monte à 52 % lorsque le contexte dans lequel les énoncés sont enregistrés est différent [HAZEN, 1973]. Quant à l'étude de BURKE ET COLEMAN, effectuée avec des examinateurs naïfs, elle met en évidence qu'aucune corrélation ne peut être tirée entre les performances d'identification et l'âge, le degré d'instruction, le degré de certitude dans la décision, le temps à disposition ou le nombre de mots cibles utilisés [BURKE ET COLEMAN, 1973].

Entre 1974 et 1979, plusieurs études évaluent les performances de la méthode spectrographique en présence de voix déguisées. HOLLIEN ET MCGLONE concluent que la probabilité d'identification correcte se situe à peine au-dessus de la chance, alors que REICH montre des réductions de performance allant de 14,2% en présence de parole lente à 35% en présence de déguisement libre [HOLLIEN ET MCGLONE, 1976 ; REICH ET AL., 1976]. Par ailleurs ces auteurs mettent en évidence des taux d'identification limités de 56,7% en l'absence de déguisement de la voix. SMRKOVSKI montre par contre que l'entraînement de l'examineur permet de réduire les taux d'erreur de manière significative [SMRKOVSKI, 1976].

TRUBY, quant à lui, dénonce le mystère qui entoure le processus de reconnaissance de formes de cette technique et pense qu'il est impossible de le définir, de le décrire et de l'évaluer [TRUBY, 1976 IN : HOLLIEN, 1990]. Finalement HOLLIEN dresse un état de la situation aux États-Unis, lors de la *Conference on Crime Countermeasures*, dans lequel il expose tous les griefs faits à la méthode spectrographique [HOLLIEN, 1977].

5.2.2.6. Analyse de l'extension de l'étude de TOSI ET AL. aux conditions forensiques réelles

5.2.2.6.1. Échantillonnage et résultats

Selon THOMAS, le processus de sélection de l'ensemble des 250 locuteurs de test dans l'étude de TOSI n'est pas aléatoire, contrairement à ce qui est prétendu. Le choix s'est porté sur un groupe de locuteurs assez homogène, qui ne peut pas être représentatif de la totalité de la population estudiantine. Les résultats obtenus ne peuvent donc pas être extrapolés à la population estudiantine dans son entier [TOSI ET AL., 1972B ; THOMAS, 1981].

5.2.2.6.2. Extension des résultats

Aucune preuve n'a été produite par TOSI pour déclarer que, dans des conditions forensiques réelles, les performances d'identification augmentent. Au contraire, d'importants facteurs non examinés, comme la faible qualité des enregistrements, pourraient les faire diminuer [SIEGEL, 1976].

5.2.2.6.3. Inférence de l'identité du locuteur

Bien que TOSI concède que d'un point de vue théorique la population potentielle est un ensemble ouvert, il affirme que dans les situations pratiques l'ensemble est un ensemble fermé. Cette conception de l'identification forensique n'est pas du tout en accord avec le concept d'identification forensique développé par TUTHILL, qui repose sur un processus d'individualisation à partir d'un ensemble de locuteurs défini selon les circonstances du cas, la population potentielle ⁴⁵ [TUTHILL, 1994]. Or les résultats de l'étude montrent que, dans une population pourtant réduite, l'augmentation du nombre de 20 à 40 locuteurs altère déjà les performances d'identification de manière significative [LASHBROOK, 1972].

⁴⁵ *supra* : 1.1.2.2. L'identification

5.2.2.6.4. Validation

CUTLER, JONES et WELCH mettent en évidence le caractère embryonnaire et expérimental de la méthode d'identification par comparaison visuelle de spectrogrammes vocaux, l'absence de fondement scientifique et de démonstration des hypothèses de base qui la sous-tendent, l'absence de démonstration de sa fiabilité, l'absence d'acceptation par la communauté scientifique, le caractère purement subjectif de l'inférence ultime de l'identité, sa valeur probante jugée incertaine et minimale et son haut potentiel trompeur à l'égard du jury [CUTLER *ET AL.*, 1972 ; JONES, 1973A ; JONES, 1973B ; WELCH, 1973]. Ils mentionnent encore les réserves qui peuvent être émises à propos des personnes impliquées dans le développement de cette méthode et concluent que ce système ne devrait être admis pour aucune tâche, ni d'exclusion ni de corroboration.

Les conclusions du rapport lui-même sont déjà équivoques :

« En général, un examinateur entraîné est clairement capable de reconnaître les spectrogrammes des mêmes mots produits par le même locuteur. De plus, lorsque des erreurs sont commises, le taux de faux négatifs (erreur de type I) est plus important que le taux de faux positifs (erreur de type II) » [LASHBROOK, 1972].

Aucun degré de validité de cette conclusion n'est précisé. Tout laisse à penser qu'il est question d'une validité générale alors qu'elle est en fait limitée à la première partie de l'étude, effectuée en laboratoire.

5.2.2.6.5. Observations du *Technical Committee on Speech Communication of the Acoustical Society of America (BOLT II)*

Dans une lettre à l'éditeur du *Journal of Acoustical Society of America*, les auteurs de BOLT I⁴⁶ mettent en évidence que certains facteurs prépondérants dans le domaine forensique n'ont pas été abordés dans l'étude de TOSI, tels les changements de l'état psychologique provoqués par le stress ou les émotions, les effets de l'environnement sonore, du système d'enregistrement, du déguisement ou de l'imitation, tous susceptibles d'augmenter la variabilité intralocuteur de la voix [BOLT *ET AL.*, 1973].

Dans leur analyse des résultats, ces mêmes auteurs soulignent aussi que, dans les conditions les plus proches des conditions forensiques présentes dans l'étude de TOSI, la comparaison de spectrogrammes non contemporains issus d'énoncés enregistrés en contexte libre et dans un ensemble ouvert aboutit à un taux d'erreur de 29 % : 5 % de fausses identifications et 24 % de fausses exclusions. Ce résultat montre que la probabilité d'erreur augmente substantiellement dans les conditions peu idéales rencontrées dans le domaine forensique. Dès lors, prétendre sans démonstration scientifique que la probabilité d'erreur sera inférieure dans les conditions forensiques que lors des expériences de laboratoire, relève d'une extrapolation fautive et abusive de la part de TOSI. THOMAS souligne aussi l'absence de tout commentaire de la part de TOSI sur le taux de non-conclusion de plus de 56% des examinateurs dans les conditions forensiques réelles, dû à la piètre qualité de l'information [THOMAS, 1981].

⁴⁶ *supra* : 5.2.1.4. Rapport du *Technical Committee on Speech Communication of the Acoustical Society of America (BOLT I)*

Finalement, vouloir démontrer la fiabilité de la méthode dans le domaine forensique par le fait que les décisions d'identification par comparaison visuelle de spectrogrammes ont toujours été corroborées par les autres éléments de l'enquête, relève de l'interprétation fallacieuse. Sans se prononcer sur la recevabilité d'une preuve obtenue par cette méthode, les auteurs de BOLT I concluent qu'il n'est scientifiquement pas possible d'évaluer la fiabilité de la méthode dans des conditions réelles [BOLT ET AL., 1973].

En réponse, une seconde lettre à l'éditeur incluant dans ses auteurs TOSI et NASH, tente de définir la communauté scientifique pertinente dans ce domaine, incluant tous les praticiens de cette nouvelle technique et excluant tous les scientifiques, pourtant respectés, ayant moins de pratique, mais plus d'objectivité scientifique [BLACK ET AL., 1974].

5.2.3. Recevabilité de la méthode spectrographique

La méthode d'identification de locuteurs par comparaison visuelle de spectrogrammes demeure l'un des moyens de preuve les plus controversés présentés devant la justice des États-Unis avec le détecteur de mensonges [REYNOLDS ET WEBER, 1979]. Preuve en est, le débat juridique nourri et ininterrompu depuis le début des années septante, dont les rebondissements illustrent à merveille les difficultés et les erreurs d'interprétation des standards de recevabilité des preuves scientifiques dans le système juridique nord-américain.

Dès 1971, TOSI et NASH unissent leurs efforts pour faire admettre cette preuve en justice et témoignent dans de nombreux cas. TOSI déposait pour démontrer la validité scientifique et la fiabilité de la méthode, ainsi que pour livrer son opinion concernant les compétences du lieutenant NASH, le décrivant comme « le meilleur examinateur de la terre ». Si la cour décrétait que la technique était admissible, NASH présentait sa méthodologie, scientifiquement démontrée par l'étude de TOSI, et concluait souvent par une variation de la phrase suivante :

« Mon opinion est que la voix inconnue et celle de l'accusé sont la même et la voix inconnue ne saurait être celle de personne d'autre ⁴⁷ » [GRUBER ET POZA, 1995].

Après l'étude de TOSI, la cour Suprême du Minnesota est la première cour civile à avoir admis ce type de preuve dans l'affaire *Trimble v Hedman* ⁴⁸. Selon elle, la technique apparaît dorénavant comme extrêmement fiable et ne justifie plus une décision de refus, comme dans les deux cas précédents [MOENSSENS ET AL., 1986]. En fait, la cour rejette le principe d'acceptation générale par la communauté scientifique pertinente en expliquant que le désaccord entre experts n'est pas rare et que c'est au magistrat instructeur de déterminer quel expert est le plus crédible [REYNOLDS ET WEBER, 1979 ; THOMAS, 1981]. L'approche développée dans *Trimble* a subséquem-

⁴⁷ [People v Law, (1974) 40 Cal App 3d 69, 114 Rptr 708, 711, 5th Dist.]

⁴⁸ [Trimble v Heldman, (1971) 291 Minn 442, 192 NW2d 432, 49 ALR3d 903]

ment été adoptée par de nombreuses cours d'État et par au moins deux cours fédérales dans *United States v Phoenix*⁴⁹ et *United States v Raymond*⁵⁰ [CUTLER ET AL., 1972].

LADEFOGED, qui avait formulé une véhémence critique de la méthode, est convaincu par les résultats de l'étude de TOSI et témoigne en cour dans le cas *Raymond*. Il prétend que la communauté scientifique est maintenant majoritairement favorable à la méthode et en informe le conseiller scientifique du président des États-Unis, Edward David Jr., dans une lettre du 24 mai 1971 en ces termes :

« Si l'on me demandait de témoigner au sujet de la validité du système, je devrais insister sur le fait que nous ne connaissons pas pour l'instant le taux d'erreur probable. Mais j'accepterais un minimum de 6 % comme estimation sommaire de la possibilité d'effectuer une fausse identification (en supposant bien sûr qu'il ne s'agisse ni de voix de femmes ni d'imitation et que l'identification ait été faite par un investigateur expérimenté et responsable) » [LADEFOGED ET VANDERSLICE, 1967].

LADEFOGED mentionne tout de même que la fiabilité de la méthode est entièrement dépendante de l'examineur, dont il propose de mesurer les performances par un indice qu'il nomme « *confusability factor* » [CUTLER ET AL., 1972].

La carrière d'examineur de spectrogrammes de NASH auprès de la police d'État du Michigan se termine cependant, après deux erreurs commises lors de dépositions en cour. La première eut lieu en 1973 dans le cas *California v Chapter*⁵¹. L'examineur se trouve face à plusieurs suspects, configuration appelée « *voice lineup* », et prétend que la voix no 4 est celle de l'accusé alors qu'il s'agit de celle du procureur du district. De plus, TOSI avait dit le jour précédent en cour qu'un examineur comparant le phonème [ɛ] de *key* avec le phonème [e] de *mate* était incompetent, ce que NASH avait précisément fait.

L'association entre TOSI et NASH continue pourtant dans *Michigan v Chaisson*⁵² en 1974. Dans ce cas, NASH rend une identification positive de l'accusé sur la base de la comparaison de douze mots cibles. La cour demande à TOSI d'examiner le cas de manière indépendante ; celui-ci conclut que les énoncés sélectionnés par NASH ne lui permettent de trouver aucune ressemblance et qu'il lui est impossible d'aboutir à une identification [POZA, 1974 IN : GRUBER ET POZA, 1995].

Malgré ces échecs, les cours fédérales continuent à admettre la comparaison visuelle de spectrogrammes comme méthode d'identification de la voix. Pour ne plus avoir à se référer au standard de *Frye*, le *Sixth Circuit* déclare, dans l'affaire *United States v Franks*⁵³, qu'acceptation générale et validité sont presque synonymes. Dans *United States v Baller*⁵⁴, le *Fourth Circuit* utilise le vecteur de la spectrographie pour rejeter le standard de *Frye* en affirmant qu'il est plus approprié

⁴⁹ [United States v Phoenix, (1971) No. 70-CR-428, S. D. Ind.]

⁵⁰ [United States v Raymond, (1972) 337 F. Supp. 641, D.D.C.]

⁵¹ [California v Chapter, (1973) Cr. 65050, Mun Ct, Marin Co, San Rafael]

⁵² [Michigan v Chaisson (1974), Ingham County Cir. Ct., No. 73'24676-FY]

⁵³ [United States v Franks, (1975) 511 F.2d 25, 33, n.12, 6th Cir., cert denied, 422 U.S. 1042]

⁵⁴ [United States v Baller, (1975) 519 F.2d 463, 465, n.1, 4th Cir. ; cert denied, 423 U.S. 1019]

d'accepter l'introduction d'éléments de preuve pertinents et de permettre au tribunal de décider lui-même de leur valeur après interrogatoire contradictoire des experts [GIANELLI ET IMWINKELRIED, 1986].

Les tribunaux des Etats sont près de rejeter ensemble le standard de *Frye*. Ils préfèrent le reformuler, afin que ses critères d'évaluation permettent la recevabilité de la méthode spectrographique. Dans l'affaire *Commonwealth v Lykus*⁵⁵, la cour Suprême du Massachusetts l'exprime de cette manière :

« Aussi limité que puisse être le nombre d'experts, les conditions requises par le standard de *Frye* sont à notre avis remplies, si le principe scientifique est accepté par ceux qui sont familiers de son usage » [REYNOLDS ET WEBER, 1979].

Une telle interprétation du concept de communauté scientifique pertinente implique que celle-ci n'est plus constituée que de personnes acquises à la méthode, nécessairement partiales et dépendantes. La cour Suprême de Californie le relève d'ailleurs dans *People v Kelly*⁵⁶ en notant que :

« Comme KERSTA avant lui, NASH avait construit sa carrière sur la fiabilité de cette technique et s'identifiait trop aux postulats de la méthode spectrographique pour juger de manière équitable et impartiale toute position scientifique opposée ».

Si la cour conserve l'interprétation reformulée du standard de *Frye*, elle juge par contre que le technicien NASH ne peut être assimilé à un scientifique [GRUBER ET POZA, 1995].

Se rapprochant de la doctrine, la cour Suprême de Pennsylvanie applique de manière stricte le standard de *Frye* dans *Commonwealth v Topa*⁵⁷ en réaffirmant que la validité de la méthode spectrographique n'a pas acquis l'acceptation générale de la communauté scientifique des sciences acoustiques. Du même point de vue, la cour Suprême du Michigan décide, dans le cas *People v Tobey*⁵⁸, que la recevabilité de la méthode spectrographique est une erreur puisque sa validité n'avait pas été établie par des experts désintéressés et impartiaux [MOENSSENS ET AL., 1986].

Si dans un premier temps la méthode spectrographique est acceptée dans le cas *Reed v State*⁵⁹, elle montre que la controverse s'est installée jusque dans le prétoire ; en effet, trois des sept juges de première instance désapprouvent publiquement la décision de recevabilité prise par la majorité et ne s'y rallient que par nécessité de collégialité. La cour d'appel intermédiaire du Maryland confirme cette décision, tout comme la cour d'appel, en reprenant l'argumentation développée dans *United States v Baller*⁶⁰. Sur la base de la critique du standard de *Frye* par MCCORMICK, elle applique le principe de pertinence de l'élément de preuve plutôt que le principe

⁵⁵ [Commonwealth v Lykus, (1975) 367 Mass. 191, 327 N.E.2d 671]

⁵⁶ [People v Kelly, (1976) 17 Cal. 3d 24, 549 P.2d 1248-1249, Cal Rptr. at 152 - 153]

⁵⁷ [Commonwealth v Topa, (1977) 471 Pa. 223, 369 A.2d 1277]

⁵⁸ [People v Tobey, (1975) 60 Mich App 420, 231 NW2d 403, 408]

⁵⁹ [Reed v State, (1978) 391 A.2d 364, Md.]

⁶⁰ [United States v Baller, (1975) 519 F.2d 463, 4th Cir. ; cert denied, 423 U.S. 1019]

d'acceptation générale [MCCORMICK, 1954 IN : GRUBER ET POZA, 1995]. En effet, selon MCCORMICK, le principe d'acceptation générale est une condition propre à la prise en compte juridique de faits scientifiques, mais n'est pas un critère de recevabilité de la preuve scientifique. Toute conclusion pertinente soutenue par un expert qualifié devrait être admise à la réserve d'autres raisons d'exclusion. Contre toute attente, la cour d'appel du Maryland rejette les décisions des cours inférieures. Consciente des critiques de conservatisme formulées à l'égard du standard de *Frye*, elle le compare à la proposition de MCCORMICK pour finalement la rejeter et décider que la controverse au sujet de la validité de la technique sous-jacente ne dépend pas des circonstances du cas et ne doit donc pas être résolue au cas par cas par des profanes.

En précisant le genre et le degré des divergences d'opinion de la communauté scientifique pertinente concernant la validité de la méthode et en concluant qu'elle ne satisfait pas au standard de *Frye*, la décision de la cour dans *Reed v State*⁶¹ marque le renouveau de l'application stricte du standard de recevabilité et du rejet de la méthode. Certaines cours comme le *Second Circuit* dans *United States v Williams*⁶² continuent cependant à l'admettre, en vertu du principe de pertinence énoncé dans les *Federal Rules of Evidence* [REYNOLDS ET WEBER, 1979 ; MOENSSENS ET AL., 1986].

5.3. Rapport du Conseil National des Sciences

En mars 1976, le FBI demande à l'Académie Nationale des Sciences des États-Unis (NAS) d'entreprendre une évaluation de la méthode d'identification de locuteurs par comparaison visuelle de spectrogrammes. Un comité de huit experts indépendants est formé. Il est composé de trois des auteurs de « BOLT I et II », BOLT, COOPER et PICKETT, de TOSI et de quatre autres scientifiques actifs dans le domaine de la parole ou de la physique, GREEN, HAMLET, MCKNIGHT et UNDERWOOD. GIANELLI et IMWINKELRIED remarquent à juste titre que, contrairement aux autres, TOSI ne doit pas être considéré comme impartial et indépendant [GIANELLI ET IMWINKELRIED 1986].

Ce comité reçoit pour mission de déterminer la validité de la technologie et sa recevabilité en cour, afin d'annihiler une controverse vieille de quinze ans qui a abouti à une position contradictoire des cours de justice de tous niveaux vis-à-vis de cette méthode. Après avoir procédé à une revue bibliographique extensive du domaine, le comité synthétise l'information sous forme d'un rapport final, comportant toutefois de nombreuses définitions et analyses nouvelles [BOLT ET AL., 1979].

5.3.1. Position du rapport sur les différents éléments de controverse

5.3.1.1. Qualité de l'information

L'existence d'une variabilité intralocuteur et interlocuteur rend l'analyse de la voix beaucoup plus proche de celle des écritures que de celle des empreintes digitales. Pour cette raison, le terme *voicegram*, vocogramme, doit être préféré à celui de « *voiceprint* », « empreinte vocale ». Il

⁶¹ [Reed v State, (1978) 391 A.2d 364, Md.]

⁶² [United States v Williams, (1978) 583 F.2d 1194, 2d Cir. ; *cert denied*, (1979) 439 U.S. 1117]

est possible que deux personnes puissent posséder des voix qui ne peuvent être discriminées dans les limites de précision à disposition. Lorsqu'un même locuteur prononce deux fois le même énoncé, les deux spectrogrammes qui en résultent ne sont jamais identiques, mais se ressemblent dès lors que les mots prononcés sont les mêmes et dans le même ordre.

5.3.1.2. Technique

Les spectrogrammes permettent d'examiner les caractéristiques suivantes : fréquences moyennes des formants des voyelles, largeur de bande des formants, périodes et striations verticales, pente des formants, durées, formes caractéristiques des fricatives et énergie entre les formants. Aucune méthode ne permet d'isoler les caractéristiques dépendantes du locuteur des caractéristiques dépendantes du contenu.

5.3.1.3. Conditions forensiques

Le recours à une analyse a lieu le plus souvent lorsqu'une ressemblance auditive frappante est constatée entre la voix d'une personne et celle constitutive de l'énoncé contesté ou lorsque la présomption de déguisement existe sur la base de cette écoute. L'examineur est donc prioritairement confronté aux cas difficiles plutôt qu'aux cas faciles. Dans les conditions forensiques, des variations peuvent être introduites par les circonstances, l'état émotionnel particulier, une façon de parler formelle ou informelle. Une classe spéciale de variation est introduite lorsqu'une personne déguise sa voix ou tente d'en imiter une autre. La détection de l'imitation est plus facile par l'observation de spectrogrammes que par l'écoute. Les échantillons de parole inconnue sont généralement obtenus lors d'enregistrements téléphoniques. La voix du locuteur peut être dégradée de multiples façons, distordue ou contaminée par du bruit dans la transmission téléphonique et lors de l'enregistrement par le système d'enregistrement.

5.3.1.4. Tâche de l'examineur

5.3.1.4.1. Analyse

La tâche d'identification de locuteurs par comparaison visuelle de spectrogrammes est pratiquée comme un art par des examineurs, dont la qualification principale est l'expérience. Elle consiste à observer et à interpréter des différences et des ressemblances. D'une certaine manière, cette tâche est analogue à la tâche de reconnaissance du scripteur par l'analyse de son écriture manuscrite. Sur la base de ces critères d'évaluation, l'examineur doit premièrement estimer un rapport de vraisemblance de l'identification par rapport à la non-identification et deuxièmement déterminer son seuil de décision.

Toutefois, selon THOMAS, les partisans de la méthode spectrographique ont admis que le processus est un art analogue à l'expertise de l'écriture manuscrite après avoir longtemps prétendu qu'il s'agissait d'une science et proposent sa recevabilité par analogie à l'expertise de l'écriture manuscrite [THOMAS, 1981]. Or cette analogie est fallacieuse, car l'expertise d'écriture peut être démontrée par l'expert et comprise par le profane, qui peut évaluer la validité des conclusions sur la base de l'interprétation des ressemblances et des divergences mises en évidence par l'expert sur les manuscrits. Par contre, la méthode spectrographique est accompagnée d'un verbiage pseudo-

scientifique propre à tromper le jury et le profane, qui ne peuvent ni discerner les ressemblances et les divergences sur les spectrogrammes, ni évaluer l'interprétation de l'expert et ses conclusions, à l'image de la réponse de NASH dans *People v Jackson*⁶³ :

« Je ne peux pas vous dire à la barre quels points de ressemblance m'ont en fait permis de conclure que c'était la même voix. Je ne sais pas ce que sont ces points de ressemblance, mais, par exemple, ici il y a un spot d'énergie qui est spécifique sur l'enregistrement de question ; il apparaît sur l'échantillon de comparaison. Il peut ne pas vous sembler le même à vous profanes non entraînés, mais pour mon œil entraîné, c'est le même. Ici il y a un spot d'énergie sur l'enregistrement de question qui est une autre particularité, qui devrait être ici sur l'enregistrement de comparaison. Ce peut ne pas être visible pour votre œil non entraîné, mais pour mon œil entraîné je sais qu'il est ici » [GRUBER ET POZA, 1995].

5.3.1.4.2. Décision

La décision d'identification est subjective, car elle est prise par l'examineur et non par le spectrographe. L'interprétation des différentes ressemblances et différences dépend de l'expérience et de l'entraînement. Elle peut provenir d'une transcription phonétique particulière qui met l'accent sur certaines ressemblances ou différences. Les différences dans la familiarité avec les distorsions provoquées par les technologies de communication et d'enregistrement font que certains examineurs sont plus habiles que d'autres dans l'interprétation des distorsions présentes dans les échantillons.

Des différences existent aussi dans l'assurance avec laquelle l'examineur approche les différentes tâches. Certains examineurs peuvent être plus réservés quant à une décision positive d'identification dans un cas de crime sérieux que dans un cas de crime mineur. Dans de nombreux cas, l'identification par la voix correspond à une partie de la preuve totale et la connaissance des autres preuves peut influencer, même inconsciemment, le point de vue de l'examineur.

5.3.2. Conclusion du rapport du Conseil National des Sciences

« Le principe de l'identification par la voix repose sur l'hypothèse que la variabilité intralocuteur est inférieure à la variabilité interlocuteur. Cependant, pour l'instant, cette hypothèse n'est confirmée ni par une théorie ni par des données scientifiques... Le comité conclut que les incertitudes techniques de la présente méthode d'identification par la voix sont si grandes que son application forensique ne doit être approchée qu'avec grande prudence. Le comité ne prend pas position pour ou contre l'utilisation forensique auditive et visuelle d'identification par la voix, mais recommande que s'il en est fait usage en cour, les limitations de la méthode soient clairement et entièrement expliquées au juge ou aux jurés » [BOLT ET AL., 1979].

Suivant les résultats de l'étude qu'il avait commandée, le FBI a confirmé sa position, prise au début des années septante⁶⁴, de ne pas proposer en cour de témoignages d'experts basés sur la

⁶³ [People v Jackson (1973) No. CR 9138, Vol. 40, Super. Ct., Riverside County, Cal.]

⁶⁴ *supra* : 5.2.1.5. Prise de position du *Federal Bureau of Investigations* (FBI)

méthode spectrographique. Par contre, la méthode a continué à être utilisée comme moyen d'investigation [KOENIG, 1980].

Ce rapport n'a malheureusement pas eu le même impact sur toutes les cours. Si certaines ont utilisé ses conclusions pour motiver un rejet de la méthode spectrographique⁶⁵, d'autres ont continué à l'accepter⁶⁶, en ignorant parfois jusqu'à l'existence de ce rapport⁶⁷ [BLACK ET AL., 1994].

5.4. Après le rapport du Conseil National des Sciences

5.4.1. La dissolution de l'IAVI

En 1980, l'IAVI est dissoute et ses membres ont pu rejoindre individuellement l'*International Association for Identification* (IAI). Suite à l'adhésion d'un nombre suffisant de membres, l'IAI a créé un sous-comité concernant l'identification de la voix, le *Voice Identification and Acoustic Analysis Subcommittee* (VIAAS). Le fonctionnement de cette société permet de l'assimiler plus à une confrérie qu'à une société savante, car elle est formée en majorité de non-scientifiques et son accès a été refusé à des scientifiques renommés sous prétexte qu'ils avaient témoigné en cour contre la méthode spectrographique [HOLLIEN, 1990]. Dès lors, les directives données par l'IAI dans ce domaine, ainsi que son programme de certification établi sur le modèle de celui de l'IAVI, sont sujets à caution [MOENSSENS ET AL., 1986].

5.4.2. L'étude du FBI

En 1986, le FBI a mené une étude sur 2000 cas d'identification de la voix répartis sur une période d'une quinzaine d'année dans le but de déterminer le taux d'erreur de la méthode dans des conditions forensiques réelles. Dans 1304 cas (65.2 %), soit aucune décision, soit une décision associée à un faible degré de confiance a été rendue, en majorité à cause de la faible qualité des enregistrements, plus rarement à cause de voix féminines à fréquence fondamentale élevée ou en présence de certains déguisements. 378 (18.9 %) décisions d'élimination et 318 (15.9 %) décisions d'identification ont été reportées alors que seulement deux (0.1 %) fausses éliminations et une (0.05 %) seule fausse identification ont été comptabilisées [KOENIG, 1986A]. L'auteur présente aussi la procédure d'analyse du FBI, qu'il compare ensuite à celle préconisée par l'IAI ; il montre que les exigences du FBI sont au moins équivalentes à celles de l'IAI, mais supérieures au niveau du nombre de mots comparés, de la qualité minimale de l'enregistrement de question et de la formation des examinateurs [KOENIG, 1986B].

GRUBER ET POZA soulignent que l'étude de KOENIG a été publiée sous forme de lettre à l'éditeur et n'a de ce fait pas été soumise à une revue par les pairs [GRUBER ET POZA, 1995]. Sa méthodologie est aussi sévèrement critiquée par SHIPP, qui relève principalement que la

⁶⁵ [Cornett v United States, (1983) 450 N.E.2d 498, Ind.]

⁶⁶ [United States v Smith, (1989) 869 F.2d 348, 7th Cir.]

⁶⁷ [United States v Maivia, (1990) F. supp. 1471, DC Hawaii 728F Supp 1471, 1471, 1478, US Dist.]

supposition qu'une décision d'identification est correcte lorsqu'elle est compatible avec l'issue du cas est fautive [SHIPP *ET AL.*, 1987]. Ces auteurs affirment avec raison qu'un critère tel qu'une décision de culpabilité ou d'innocence n'est pas suffisant pour établir la rectitude des décisions d'identification.

Cette étude illustre aussi l'influence prépondérante des résultats de la méthode spectrographique sur l'issue de nombreux cas, alors qu'elle est officiellement utilisée uniquement à des fins d'enquête. Cette analyse laisse à penser que la position officielle du FBI permet à l'agence d'utiliser la méthode en évitant soigneusement toute polémique concernant la recevabilité, plutôt qu'à garantir au justiciable une investigation sur la base de méthodes acceptables.

Dans une première réponse, KOENIG *ET AL.* ne se défendent pas sur le fond, mais affirment qu'ils ne considèrent pas une méthode entachée d'un taux d'erreur supérieur à 0.5% au même titre que les empreintes digitales, qu'ils considèrent comme exactes à 100%. Cette raison motive, selon eux, le fait que la méthode spectrographique n'est utilisée que pour l'investigation [KOENIG *ET AL.*, 1987].

Dans une seconde réponse, MELVIN explique qu'une probabilité plus grande que 95% est considérée comme une certitude dans toute science expérimentale [MELVIN *ET AL.*, 1988]. Cet argument illustre la confusion faite entre degré de signification associé à une observation statistique et degré de certitude associé à une décision. De plus, la démarche utilisée pour la confirmation des hypothèses va à l'encontre de la règle de falsifiabilité énoncée par POPPER, qui vise à évaluer la validité scientifique d'une hypothèse en définissant les conditions qui permettent de la réfuter [POPPER, 1973]. Cette polémique est encore alimentée par TOSI dans un éditorial du *Journal of Forensic Identification*, mais n'aborde malheureusement jamais les problèmes de fond [TOSI, 1990].

5.4.3. Les standards de l'IAI

En 1991, le sous-comité VIAAS de l'IAI publie des standards pour la comparaison des voix [VIAAS, 1992]. Ils n'ont pas de valeur légale, car ils ne lient que les examinateurs certifiés officiellement par l'IAI, mais ont l'avantage d'explicitier la méthodologie et de mettre à jour ses faiblesses intrinsèques, potentiellement impossibles à résoudre.

En résumé, l'examineur doit être adéquatement formé, entraîné et qualifié (1), l'élément de preuve doit être manipulé avec précaution (2), les échantillons doivent être soigneusement choisis en vue de la comparaison (3), des échantillons de comparaison doivent être soigneusement préparés (4). Un examen préliminaire (5) doit permettre d'évaluer la qualité des éléments de preuve et de déterminer si l'analyse peut être effectuée, par méthode auditive et spectrographique (6). L'examineur doit aboutir à l'une des sept conclusions possibles : identification, identification probable, identification possible, résultat inconcluant, exclusion possible, exclusion probable, exclusion, (7) et peut parfois demander un second avis. Le travail doit être soigneusement documenté et le rapport rédigé sous une forme standardisée. Finalement, l'IAI précise encore

qu'elle n'approuve l'usage d'aucune autre méthode d'identification de voix que celle stipulée dans ses standards.

Durant l'examen préliminaire, l'examineur doit s'assurer que les enregistrements inconnus et de comparaison sont originaux. Comme la méthode est dépendante du texte, les enregistrements de parole inconnue et de comparaison doivent comprendre au moins dix mots correspondants et des passages contenant au moins trois mots correspondants consécutifs. Les échantillons doivent être de haute qualité, sans déguisement, sans excès de distorsion, sans interférences causées par de la parole ou du bruit et sans excès de variation des systèmes de transmission, d'enregistrement ou d'autres différences pouvant détériorer notablement les caractéristiques auditives et spectrales. Finalement le signal doit posséder une bande passante et un rapport signal sur bruit suffisants.

Ces critères découlent directement des remarques formulées par TURNER dans la conclusion de l'étude de TOSI [TURNER *ET AL.*, 1972]. Leur faiblesse résulte du fait que leur évaluation est principalement subjective et que certains sont intrinsèquement incontrôlables, comme le déguisement. D'autres ne sont pas maîtrisés par l'examineur de spectrogrammes, comme la qualité d'enregistrement. La définition de tels critères va à l'encontre de la réalité forensique, puisque dans l'étude de TOSI elle-même, on a dû renoncer à toute analyse dans 57% des cas, pour cause de qualité insuffisante des échantillons.

5.4.4. L'arrêt Daubert

Selon la *Federal Rule of Evidence* 901(b)(5)⁶⁸, un témoignage reposant sur l'identification d'un locuteur est admissible. Par contre, en tant que preuve scientifique, la reconnaissance de locuteurs par comparaison visuelle de spectrogrammes est soumise au nouveau standard dit « de validité », énoncé par la cour Suprême des États-Unis dans l'arrêt Daubert⁶⁹. Cette décision représente un tournant dans la manière d'aborder un moyen de preuve scientifique nouveau ou controversé.

Si l'acceptation générale par la communauté scientifique pertinente demeure un facteur important, ce n'est plus une considération fondamentale ni un motif de rejet. Par contre, selon l'interprétation de la *Federal Rule of Evidence* 104(a) dans l'arrêt Daubert, la cour doit partir du principe que le raisonnement et la méthodologie qui sous-tendent le témoignage sont scientifiquement valides et peuvent être appliqués intégralement dans le cas d'espèce. Ensuite la cour doit évaluer dans quelle mesure la preuve scientifique présentée satisfait aux critères énoncés dans la *Federal Rule of Evidence* 702 [BLACK *ET AL.*, 1994].

Une exigence préliminaire stipule que le témoignage doit être prononcé par un expert en sciences forensiques. Cependant, l'absence de définition de cette notion laisse ouvertes tant la question des qualifications requises pour témoigner que la possibilité pour la cour de récuser l'expert, comme dans l'affaire *People v Kelly*⁷⁰, régie par l'ancien standard de recevabilité.

⁶⁸ *supra* : 3.3. Exigences légales en matière de preuve scientifique

⁶⁹ [Daubert v Merrell Dow Pharmaceuticals, (1993) US, 125 L Ed 2, 469]

⁷⁰ [People v Kelly, (1976) 17 Cal. 3d 24, 549 P.2d 1248-1249, Cal Rptr. at 152-153]

Le premier critère concerne la falsifiabilité de la méthode, sa capacité à être testée et le fait d'avoir été testée dans des conditions forensiques réalistes. Manifestement, la méthode spectrographique ne satisfait pas à ce critère. Bien que la méthode puisse être testée et qu'elle ait été abondamment testée, les résultats obtenus dans des conditions forensiques montrent qu'elle n'est guère utilisable dans ces cas-là. D'autre part, s'il est possible de tester la perception sensorielle humaine dans des conditions définies, le fait que l'examineur connaisse les circonstances de l'affaire, analyse les conséquences de sa décision, voire subisse l'influence de son jugement personnel ou celui de son entourage, rend la méthode difficilement falsifiable.

Le deuxième critère exige que la méthode ait fait l'objet d'un examen attentif et de publications, alors que le quatrième exige qu'elle soit généralement acceptée dans la communauté scientifique pertinente. La controverse qui a existé et qui se perpétue autour de cette méthode ne lui permet certainement pas de satisfaire à ces deux critères [GRUBER ET POZA, 1995].

Le troisième critère mentionne que la méthode doit avoir un taux d'erreur connu ou potentiel dans l'application. Bien que la Cour Suprême des États-Unis mentionne la comparaison visuelle de spectrogrammes comme méthode dont le taux d'erreur est connu dans l'arrêt Daubert, la controverse reste vive sur le fait que les taux d'erreur lors des expérimentations sont considérablement réduits par rapport aux taux d'erreur existant dans des conditions forensiques réelles [BLACK ET AL., 1994]. Dans *United States v Smith*⁷¹, la cour mentionne d'ailleurs que les taux d'erreur varient de 0 à 83 % selon les évaluations.

Le cinquième critère indique que le témoignage doit être basé sur des faits ou des données dignes de confiance pour les experts du domaine. Le fait que les tentatives de démonstration de la méthode spectrographique reposent essentiellement sur des données souvent considérées comme incomplètes et sur des faits discutables et critiqués par nombres d'experts reconnus dans leurs domaines tels les scientifiques de BOLT I et II, HECKER ou les juristes THOMAS ou BLACK, pour ne citer qu'eux, montre clairement que la méthode spectrographique ne satisfait pas à ce critère non plus.

Le dernier critère exige que la valeur probante de la méthode ne soit pas supplantée par les dangers d'un préjudice injuste, la confusion des conclusions ou l'induction en erreur du jury. Or, l'argument développé par les défenseurs de la méthode spectrographique allègue que seuls les praticiens de la comparaison visuelle de spectrogrammes sont à même d'en comprendre précisément le fonctionnement, alors que des scientifiques reconnus ne le peuvent pas, malgré un bagage théorique plus important. Selon TRUBY, l'impossibilité pour quiconque de définir, de décrire et d'évaluer le processus de reconnaissance de formes utilisé dans cette technique ou encore les dépositions de NASH dans *People v Jackson*⁷², qualifiées de « verbiage pseudo-scientifique propre à tromper le jury et le profane » par THOMAS, ne plaide pas en faveur de sa clarté et de sa

⁷¹ [United States v Smith, (1989) 869 F.2d 348, 7th Cir.]

⁷² [People v Jackson (1973) No. CR 9138, Vol. 40, Super. Ct., Riverside County, Cal.]

compréhension par le profane, mais témoigne plutôt de la part d'obscurantisme qui l'entoure [THOMAS, 1981 ; GIANELLI ET IMWINKELRIED, 1986 ; TRUBY, 1976 IN : HOLLIEN, 1990].

5.5. La méthode spectrographique dans le reste du monde

En Europe, l'annonce de l'existence de la méthode de KERSTA soulève un enthousiasme béat, dans un premier temps. Aucun essai n'est réalisé en Europe puisqu'aucun spectrographe sonore n'est disponible. Quelques auteurs, dont les compétences dans le domaine particulier de l'identification de locuteurs sont discutables, se contentent de répéter sans aucun sens critique les propos de KERSTA, certains d'avoir affaire à une découverte majeure dans le domaine des sciences forensiques [MARTIN, 1967 ; ROTHER 1967 ; HABERSBRUNNER ET AL., 1968].

Cependant, en Allemagne a lieu une tentative infructueuse de vérification des hypothèses de KERSTA par ENDRESS⁷³. Ceci soulève une vague de scepticisme, notamment en France, de la part du docteur Tomatis, fondateur de l'Institut qui porte son nom, du professeur Vallancien, responsable de l'Institut français de la Voix, ou encore de l'ingénieur des Télécommunications MAMOUX, expert près la cour d'appel de Paris [ENDRESS ET AL., 1971 ; MAMOUX, 1971 ; BLOCK, 1975]. Ces prises de position conduisent les principaux pays européens membres de l'Interpol à adopter une attitude circonspecte vis-à-vis de cette technique. Selon BLOCK, seule la Suisse fait exception à la règle [BLOCK, 1975].

L'attitude la plus extrême et la moins clairvoyante est certainement celle adoptée par la Roumanie, qui possède une méthode spectrographique d'identification de personnes par la voix très inspirée de celle de KERSTA et brevetée en 1972 [ANGHELESCU, 1974]. Cette méthode permet selon son auteur :

1. La détermination du sexe de la personne.
2. L'établissement de certaines données sur l'identité du sujet et de certaines maladies dont il peut être atteint.
3. L'identification du sujet d'après la voix et la phonation utilisées dans le cadre de conversations courantes.
4. La découverte et la démonstration des falsifications de la voix et de la phonation.

L'approche roumaine reflète sans aucun doute l'esprit du régime politique d'alors, car le même auteur, devenu directeur de l'Institut de Criminalistique de l'Inspectorat général de la milice de Roumanie, insiste dans un second article publié en 1985 dans la Revue Internationale de Police Criminelle (RIPC) [ANGHELESCU, 1985] :

« Les investigations ont démontré que l'hypothèse selon laquelle la voix et le parler permettent l'identification du sujet est scientifiquement fondée puisque la voix reste stable au cours de la

⁷³ *supra* : 5.2.1.3. Prise de position de la communauté scientifique et juridique sur l'étude de KERSTA

vie adulte jusqu'à la vieillesse ; toutes les modifications vocales consécutives au vieillissement deviennent à leur tour des particularités concourant à l'identification.

... Depuis 1972, le laboratoire de phonocriminalistique de notre institut a effectué un nombre appréciable d'expertises ayant pour objet l'identification des personnes d'après la voix et le parler, expertises dont les résultats ont été confirmés dans la proportion de 99 % par les tribunaux.

... Depuis quelque temps, la criminalistique s'est emparée, pour les mettre au service de la justice et de la vérité, des plus récentes réalisations techniques et scientifiques contemporaines, au nombre desquelles figure, à côté de l'ordinateur, le laser. On en est donc venu à examiner les diagrammes en les éclairant à l'aide d'une source laser dans un système de filtrage optique, ce qui permet d'obtenir un spectre de Fourier qui synthétise toutes les caractéristiques de la voix transcrites sur les sonagrammes et qui offre la possibilité d'un examen comparatif intégral de ces derniers. De plus, on assure ainsi la totale objectivité de ce genre nouveau d'expertise criminalistique ».

Actuellement il est très difficile de connaître le degré d'utilisation de la comparaison de spectrogrammes vocaux en Europe. La représentation spectrographique fait partie de la méthode phonétique acoustique développée par les experts phonéticiens, mais la question de son application à des buts de comparaison, telle qu'elle est pratiquée aux États-Unis, n'est pas résolue puisqu'aucun consensus n'existe entre les experts et qu'en conséquence aucune méthodologie commune et explicite n'a été publiée. En l'absence d'une telle publication et à cause de la discrétion des experts sur cette question, tout laisse à penser que son utilisation dépend de la méthodologie propre à l'expert et des circonstances du cas.

5.6. Conclusion

D'un point de vue scientifique, il semble aujourd'hui acquis que la méthode d'identification de locuteurs par comparaison visuelle de spectrogrammes vocaux ne peut pas être considérée comme valide et qu'elle n'est pas utilisable dans le domaine forensique.

L'intérêt principal de l'étude de l'approche spectrographique réside dans le fait que la controverse suscitée depuis 1962 a forcé les différents acteurs du monde judiciaire nord-américain à réfléchir à la notion de validité scientifique et à expliciter les critères de recevabilité de la preuve scientifique. Elle a aussi contribué à faire prendre conscience aux experts forensiques que l'utilisation de toute méthode est subordonnée à une compréhension intégrale et complète des principes scientifiques et du jeu d'inférences qui la sous-tendent.

L'existence de ce regard critique, moteur du progrès scientifique, est à mettre à l'actif de la procédure accusatoire qui favorise le débat, malheureusement parfois jusqu'à l'excès. A l'inverse, le système inquisitoire en vigueur en Suisse occulte tout débat sur ces sujets pourtant cruciaux, ce qui limite fortement la remise en cause des experts et une évolution conjointe de leurs méthodes à celle du progrès scientifique.

Dans cette controverse, il est peu enthousiasmant de noter la propension des scientifiques à se pencher sur de faux problèmes, comme de vouloir évaluer la capacité des experts au lieu de

démontrer la validité des hypothèses et de la technique, tout comme de constater l'inaptitude du système juridique nord-américain à analyser de manière correcte la validité de cette méthode, cataloguée *a priori* de scientifique.

Finalement, aujourd'hui encore, malgré de multiples prises de position de scientifiques renommés, un rapport de l'Académie Nationale des Sciences des États-Unis et un certain nombre d'arrêts de la Cour Suprême dans la ligne de l'arrêt Daubert, le système juridique nord-américain est malheureusement toujours incapable de se déterminer de manière définitive.

VI. APPROCHE AUTOMATIQUE

6.1. Introduction

6.1.1. Définition

« La reconnaissance automatique de locuteurs est l'étude de la capacité de l'outil informatique à procéder à la reconnaissance de personnes à partir d'une donnée biométrique variable, la voix, sur la base de méthodes exploitant la théorie de l'information, la reconnaissance automatique de formes et l'intelligence artificielle perceptive » [BUNGE, 1991].

Cette technologie ouvre potentiellement la voie à plusieurs applications commerciales, comme le contrôle d'accès, physique ou à de l'information, ainsi qu'aux deux applications forensiques que sont la reconnaissance de locuteurs à partir d'enregistrements présentés comme indices et la surveillance lors d'incarcération à domicile [DODDINGTON, 1985 ; BOVES, 1998].

6.1.2. Historique

Les premières méthodes de reconnaissance automatique de locuteurs ont été développées à partir du début des années 1960 [PRUZANSKI, 1963 ; PRUZANSKI ET MATTHEWS, 1964 ; LI ET AL., 1966 ; RAMISHVILI, 1966 ; ATAL 1968 ; LUCK, 1969]. Dans les années 1970, les recherches en vue d'applications commerciales émanent essentiellement des centres de recherche liés à de grands constructeurs informatiques tels que *International Business Machines*[®] (IBM) [DAS ET MOHN, 1971] et *Texas Instruments*[®] [DODDINGTON, 1976], ou à des compagnies de télécommunication comme *American Telephone and Telegraphs*[®] (AT&T) [BRICKER ET AL., 1971 ; SAMBUR, 1975 ; ROSENBERG, 1976A] ou *Nippon Telephones and Telegraphs*[®] (NTT) [FURUI, 1981B] [CAPPE, 1995].

Des recherches en vue d'applications forensiques sont aussi entreprises, comme le *Semi Automatic Speaker Identification System* (SASIS) développé par le *Stanford Research Institute* (SRI) puis *Rockwell International*[®], ou le projet *AUtomatic Recognition Of Speakers by computers* (AUROS) conçu en Allemagne chez *Philips GmbH*, puis développé au BKA [BECKER ET AL., 1973 ; PAUL ET AL., 1975 ; BUNGE, 1977].

A partir des années 1980, le développement de l'informatique et celui de nouvelles méthodes ont dynamisé la recherche dans le domaine de la reconnaissance de locuteurs et actuellement de nombreux organismes à travers le monde poursuivent des recherches dans ce domaine, comme le montre la participation d'une quinzaine de laboratoires à la dernière évaluation des algorithmes de reconnaissance de locuteurs proposée par le *National Institute of Standards* (NIST) des États-Unis [PRZYBOCKI ET MARTIN, 1998].

6.2. Analyse du signal de parole

6.2.1. Principes

Dans le domaine de la reconnaissance de locuteurs, le traitement du signal vocal a pour but de fournir une représentation moins redondante de la parole que celle obtenue par l'onde temporelle, tout en permettant une extraction précise des caractéristiques significatives pour la reconnaissance de la parole ou de locuteurs. Cette analyse s'appuie soit sur l'extraction et la reconnaissance automatique d'événements acoustiques correspondant aux éléments phonétiques, soit sur la variabilité implicite du signal de parole en fonction du locuteur [ROSENBERG, 1976B ; MELLA, 1992]. Les principaux problèmes posés en traitement automatique de la parole proviennent de la dualité source-conduit de l'appareil phonatoire et de la grande dynamique et de la variété des voix.

6.2.1.1. Analyse phonétique-acoustique

Les premières recherches se sont concentrées sur une analyse phonétique-acoustique du signal de parole, dans le but de découvrir les caractéristiques temporelles et spectrales les plus dépendantes du locuteur. L'extraction de caractéristiques phonétiques-acoustiques de manière auditive, comme celle pratiquée par les experts-phonéticiens⁷⁴, est aisée car l'oreille humaine est excellente dans la discrimination des signaux de parole pertinents dans des milieux fortement bruités, plus particulièrement lorsque ce bruit est composé de parole. Cet effet, appelé *cocktail party effect*, pose par contre beaucoup de difficultés aux algorithmes utilisés pour l'analyse automatique du signal de parole, difficultés de segmentation automatique des phonèmes dans les applications dépendantes du texte et d'identification automatique des phonèmes dans les applications indépendantes du texte, comme le montrent les expériences de DAS ET MOHN et celles de HAIR ET REKIETA [DAS ET MOHN, 1971 ; HAIR ET REKIETA, 1972 ; INGRAM, 1995].

Même la méthode récente développée par NEWMAN, où la segmentation des phonèmes repose sur une reconnaissance de parole indépendante du locuteur dans un vocabulaire étendu, ne permet pas de surpasser les méthodes basées sur une analyse de la variabilité implicite du signal de parole [NEWMAN ET AL., 1996 IN : FURUI, 1997]. Ces obstacles ont conduit les chercheurs à se tourner vers l'extraction de caractéristiques dynamiques et statistiques, mesurables tout au long du signal de parole [FURUI, 1981B ; O'SHAUGNESSY, 1986]. Cette évolution fait suite à ce qui s'est passé dans le domaine de la reconnaissance de la parole où l'approche acoustique-phonétique, basée sur l'extraction et la reconnaissance explicite des événements acoustiques correspondant aux éléments phonétiques, s'est avérée moins efficace, dans le cadre d'applications pratiques, que les méthodes de type reconnaissance de formes [RABINER ET JUANG, 1993].

⁷⁴ *supra* : 4.4.2. L'approche phonétique acoustique

6.2.1.2. Analyse de la variabilité implicite du signal en fonction du locuteur

L'analyse de la variabilité implicite du signal en fonction du locuteur a conduit au développement de trois catégories d'algorithmes : les analyses à court terme, temporelles, spectrales et spectro-temporelles ; les méthodes fondées sur la déconvolution source-conduit, homomorphiques ou basées sur la prédiction linéaire ainsi que les méthodes fondées sur un modèle d'audition, comme les bancs de filtres. Les méthodes d'analyse à court terme, temporelles, spectrales ou spectro-temporelles reposent sur une description mathématique rigoureuse, mais ne se réfèrent pas toujours à un modèle de production ou de perception. Les méthodes fondées sur la déconvolution source-conduit n'ont pas ce défaut, mais reposent sur un modèle de production souvent imprécis. Finalement, les méthodes fondées sur un modèle d'audition ne garantissent pas d'adéquation entre ce qui est perçu et les résultats de l'analyse du fait de l'imbrication, chez l'humain, des niveaux d'interprétation acoustique et linguistique [DRYGAJLO, 1999].

6.2.2. Approches primaires

6.2.2.1. Analyse temporelle

Les analyses les plus simples consistent à mesurer l'énergie, le taux de passage par zéro et la fonction d'autocorrélation à court terme du signal.

6.2.2.1.1. Énergie

L'évolution à court terme de l'énergie du signal indique la succession des voyelles, très énergétiques, et des consonnes, de moindre énergie. LUMMIS a utilisé la valeur absolue de la différence entre les énergies de deux échantillons pour procéder à leur alignement temporel en mode dépendant du texte [LUMMIS, 1973].

6.2.2.1.2. Taux de passage par zéro

Pour un signal numérisé, il y a passage par zéro dans la représentation temporelle lorsque deux échantillons successifs sont de signes opposés. Le comptage et le tracé d'histogrammes des passages par zéro du signal traduisent, bien que grossièrement, le contenu spectral. Les valeurs du taux de passage par zéro sont normalement plus élevées pour les sons non voisés que pour les sons voisés. Le taux de passage par zéro à long terme présente une répartition sensiblement gaussienne avec une moyenne de l'ordre de $4,9 \text{ ms}^{-1}$ pour les sons non voisés et de $1,4 \text{ ms}^{-1}$ pour les sons voisés ; ces deux répartitions se recouvrent partiellement.

GUBRYNOWICZ, LIN ET PILLAY, ainsi que BASZTURA ET JURKIEWICZ ont montré l'efficacité de l'analyse à court terme du taux de passage par zéro en mode dépendant du texte, alors que BASZTURA ET MAJEWSKI ont montré l'efficacité de l'analyse à long terme de cette caractéristique en mode indépendant du texte, sur des échantillons de parole de 30 s à 40 s [GUBRYNOWICZ, 1973 ; LIN ET PILLAY, 1976 ; BASZTURA ET JURKIEWICZ, 1978 ; BASZTURA ET MAJEWSKI, 1978]. GOPALAN ET MAHIL ont montré qu'associé à la mesure de l'énergie à court terme, le taux de passage par zéro est efficace pour l'analyse d'échantillons de très courte durée [GOPALAN ET MAHIL, 1991].

6.2.2.1.3. Fonction d'autocorrélation

La variation à court terme de la fréquence fondamentale, ou contour de F_0 , connaît d'importantes variations interlocuteurs. Comme plusieurs méthodes permettent de l'extraire automatiquement du signal de façon fiable même en cas de rapport signal sur bruit faible, cette caractéristique a été largement utilisée pour la reconnaissance automatique de locuteurs [ATAL, 1976 ; CORSI, 1982].

Dans le domaine temporel, le calcul de la fonction d'autocorrélation à court terme de la forme d'onde filtrée permet d'extraire la périodicité, en déterminant le degré de similarité entre deux courbes. La fréquence fondamentale peut alors être estimée en cherchant le 2^{ème} pic le plus important de la fonction d'autocorrélation. L'efficacité de cette approche temporelle d'analyse de F_0 pour la reconnaissance automatique de locuteurs a notamment été montrée par ATAL. Dès 1968, il a obtenu un taux d'identification de 97% en mode dépendant du texte, sur une base de données de dix locutrices, en analysant l'évolution dynamique de la fréquence fondamentale sur des échantillons de parole normalisés dans le domaine temporel [ATAL, 1968].

Par l'analyse de la distribution statistique de la moyenne à court terme de la fréquence fondamentale, STEFFEN-BATOG estime qu'il est possible de différencier une cinquantaine de locuteurs masculins à partir d'échantillons de 10 à 50 s de texte lu, mais JASSEM conclut qu'en cas d'application forensique, une telle méthode est plus indiquée pour une classification préliminaire des suspects que pour leur identification [STEFFEN-BATOG ET AL., 1970 ; JASSEM ET AL., 1973]. Si la valeur de la fréquence fondamentale est moyennée sur une période de temps suffisamment longue, elle est relativement constante dans le temps et indépendante du contexte linguistique [HORII, 1975].

Grâce à la fonction d'autocorrélation, ATKINSON a montré de grandes différences entre la fréquence fondamentale des enfants, des femmes et des hommes, qui est perçue de manière correcte auditivement, mais il a aussi mis en évidence que les variations interlocuteur et intralocuteur de F_0 sont malheureusement concurrentes [ATKINSON, 1976].

L'observation de la supériorité des caractéristiques spectrales a réduit l'intérêt des chercheurs pour les caractéristiques temporelles comme la fréquence fondamentale, la durée de phonation, le débit de parole ou le développement de fonctions temporelles statistiques, pourtant potentiellement porteuses d'informations dépendantes du locuteur [GROSJEAN ET DESCHAMPS, 1972 ; DODDINGTON, 1985 ; O'SHAUGNESSY, 1986].

6.2.2.2. Analyse spectrale

Les méthodes spectrales sont fondées sur une décomposition fréquentielle du signal sans connaissance *a priori* de sa structure fine. La seule hypothèse mise en jeu concerne le choix des fonctions sur la base desquelles le signal est décomposé : en fonctions sinusoïdales pour la transformée de Fourier, en fonctions créneaux pour la transformée de Walsh-Hadamard ou encore par le choix des caractéristiques des filtres pour une analyse en banc de filtres. Dans une certaine mesure, ce choix peut être considéré comme dépendant de la structure de la parole.

6.2.2.2.1. Analyse du spectre à court terme par banc de filtres

L'analyse du spectre de puissance à court terme a d'abord été obtenue par le passage de la parole à travers des bancs de filtres. L'énergie à la sortie de chaque filtre passe-bande fournit une bonne estimation du spectre à court terme de la fréquence centrale du filtre [ATAL, 1976]. L'extraction de caractéristiques dépendantes du locuteur par ce moyen a été initiée par PRUZANSKI ET MATHEWS, ainsi que par LI [PRUZANSKI, 1963 ; PRUZANSKI ET MATHEWS, 1964 ; LI ET AL., 1966 ; BRICKER ET AL., 1971]. Un système de contrôle d'accès, basé sur un banc de filtres à 14 canaux, a été développé par DODDINGTON, pour le centre de calcul de *Texas Instruments*[®]. Il est opérationnel, mais il a permis de montrer que la dynamique vocale est facile à imiter par un imitateur, surtout en cas de répétition immédiate de l'énoncé entendu [DODDINGTON, 1979].

6.2.2.2.2. Analyse du spectre à court terme par transformée de Fourier

La correspondance numérique de la transformée de Fourier est la transformée de Fourier discrète. En principe, le concept de la transformée de Fourier discrète ne s'applique qu'à un signal stationnaire de durée limitée. Comme le signal vocal est essentiellement non stationnaire, cette notion est remplacée par celle de transformée de Fourier à court terme.

Les propriétés de la transformée de Fourier à court terme dépendent beaucoup du choix de la fonction fenêtre. D'une part la longueur de cette fenêtre doit être suffisante pour assurer une bonne résolution et d'autre part elle doit être suffisamment limitée pour suivre fidèlement l'évolution du spectre vocal dans le temps. Ces deux exigences sont contradictoires. Pour suivre au mieux les transitions de la parole, il est nécessaire de prendre des fenêtres temporelles avec recouvrement. Mais il demeure malgré tout un effet de lissage temporel dû à la longueur de cette fenêtre ainsi qu'une distorsion du spectre dépendant du type de fenêtre utilisé [DRYGAJLO, 1999]. BLACKMAN ET TUKEY proposent l'utilisation d'une fonction fenêtre de type « *Hamming* » et une durée de 25 ms pour l'extraction du spectre de puissance à court terme [BLACKMAN ET TUKEY, 1959 IN : ATAL, 1976].

Un intérêt particulier a été porté au spectre à court terme des voyelles et des nasales explicitement extraites du signal, à cause de leur grande dépendance au locuteur et de leur identification relativement facile dans le signal de parole [LI ET HUGUES, 1974 ; SAMBUR, 1975 ; SU ET AL. 1979]. Cependant, la complexité de la procédure d'extraction automatique de ces éléments phonétiques, comme celle proposée par DAS ET MOHN, a conduit les auteurs à préférer le spectre moyen à long terme au spectre à court terme, pour sa facilité de mesure et la possibilité de l'utiliser dans un mode indépendant du texte [DAS ET MOHN, 1971 ; O'SHAUGNESSY, 1986].

6.2.2.2.3. Spectrogrammes numériques par transformée de Fourier rapide

La transformée de Fourier à court terme est souvent utilisée pour confectionner des spectrogrammes numériques. D'une part le volume de calcul qu'elle implique n'est pas trop important, grâce à l'utilisation de l'algorithme de calcul rapide *Fast Fourier Transform* (FFT), et d'autre part l'image obtenue est proche de celle du spectrogramme analogique.

Pour cette application le choix du type de la fenêtre n'est pas déterminant, car les premiers formants sont en général assez nets si la fenêtre est de longueur suffisante. La longueur de la

fenêtre détermine l'observation de la structure harmonique (Figure VI.1) ou formantique (Figure VI.2) d'un son voisé.

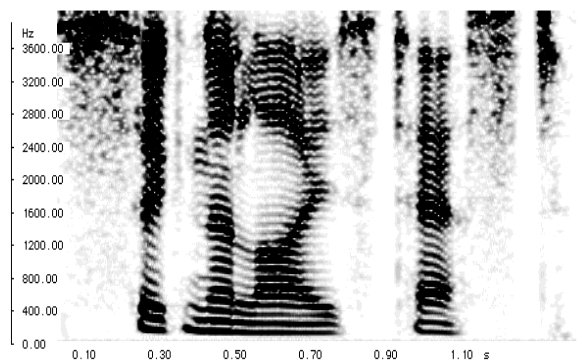


Figure VI.1. Spectrogramme à bande étroite (25ms - 40Hz) de l'énoncé «Signalyze test»

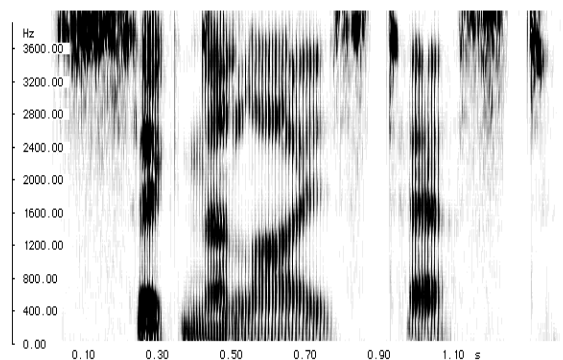


Figure VI.2. Spectrogramme à bande large (8 ms - 125Hz) de l'énoncé «Signalyze test»

Le spectrogramme est dit à « bande étroite » ou à « large bande », par analogie à la bande passante des filtres utilisés par le spectrographe dans le domaine analogique. Dans la parole continue, l'analyse en bande étroite du spectre à court terme permet d'apprécier la structure harmonique des sons voisés et l'analyse en bande large leur structure formantique. Le calcul de l'intensité s'effectue en décibels sur le module du spectre et la phase n'est pas exploitée dans cette analyse.

6.2.2.3. Sélection et exploitation des caractéristiques

6.2.2.3.1. Analyse à court terme

L'analyse à court terme consiste à calculer un ensemble de coefficients acoustiques à des intervalles de temps réguliers nommés trames, compris entre 10 et 20 ms. Ces coefficients sont déterminés à partir de fenêtres de signal représentant 20 à 40 ms de parole et l'ensemble des coefficients résultant de ces mesures constitue une trame acoustique. La pondération du signal par une fonction fenêtre permet de tenir compte du caractère non stationnaire du signal [DRYGAJLO, 1999]. En mode de reconnaissance dépendant du texte, les caractéristiques sélectionnées sont mesurées à différents instants déterminés. L'ensemble de ces mesures permet de décrire l'évolution temporelle dynamique de chacune des caractéristiques considérées sous forme d'un profil (*contour*) [O'SHAUGNESSY, 1987 ; CAPPE, 1995]. La comparaison de deux profils est réalisée par le calcul d'une distance moyenne, après un alignement temporel servant à corriger les décalages temporels pouvant exister entre deux énoncés d'un même texte [ROSENBERG, 1976B].

6.2.2.3.2. Analyse à long terme

L'indépendance du texte s'obtient le plus souvent en s'intéressant à la densité de probabilité ou à la moyenne d'une suite de coefficients à court terme extraits d'une locution, dont l'estimation se fait sur un temps suffisamment long pour pouvoir modéliser le comportement global du

locuteur [THEVENAZ, 1993]. Seules ces informations statistiques à long terme sont utilisées pour la reconnaissance de locuteurs dans ce cas.

6.2.2.3.3. Mesure d'efficacité

La mesure d'efficacité des caractéristiques le plus souvent décrite est une technique d'analyse de variance basée sur le discriminant de Fischer, nommée *F-ratio*, qui permet de sélectionner les caractéristiques dont la variabilité intralocuteur est faible et dont la variabilité interlocuteur est forte [BRICKER, 1971 *ET AL.* ; WOLF, 1972]. Cette procédure de sélection consiste à extraire le sous-ensemble des caractéristiques possédant le *F-ratio* maximal à partir d'un grand nombre de caractéristiques.

6.2.2.4. Conclusion

Les transformations spectro-temporelles considèrent le signal de parole sous forme de fenêtres successives et une transformation s'opère sur chaque fenêtre. La durée et la forme des fenêtres sont ajustées pour favoriser les interprétations recherchées. Le défaut de ces méthodes réside dans l'intermodulation source – conduit, qui rend difficile la mesure de la fréquence fondamentale et la mesure des formants, caractéristiques de la source et du conduit vocal [DRYGAJLO, 1999].

6.2.3. Approches actuelles

Contrairement aux méthodes précédentes, la prédiction linéaire et l'analyse homomorphique sont fondées sur une connaissance des mécanismes de production de la parole.

6.2.3.1. Prédiction linéaire

6.2.3.1.1. Principe

La méthode de prédiction linéaire, *Linear Predictive Coder* (LPC), aussi appelée modélisation autorégressive, est utilisée en premier lieu pour le codage du signal de parole, mais elle permet une caractérisation de l'enveloppe spectrale de ce signal dans le domaine temporel [SCHAFER ET RABINER, 1975]. Elle repose sur une modélisation paramétrique temporelle du signal de parole et se fonde sur l'observation que chaque nouvel échantillon du signal de parole ne constitue pas une innovation pure, mais qu'il est fortement corrélé aux échantillons qui l'ont précédés. Comme cet échantillon peut être prédit à partir des échantillons passés, il suffit, en principe, de calculer les coefficients et l'erreur de prédiction en utilisant la fonction d'autocorrélation [DRYGAJLO, 1999]. Le coût opératoire suffisamment faible pour permettre l'application en temps réel de la méthode de prédiction linéaire à des fins de codage ou de reconnaissance de parole en font une méthode répandue [THEVENAZ, 1993 ; DRYGAJLO, 1999].

6.2.3.1.2. Minimisation de l'énergie résiduelle de prédiction

Le calcul de l'erreur de prédiction se fonde sur les connaissances du modèle de production de parole et suppose que ce modèle de production est linéaire et que sa fonction de transfert ne comporte que des pôles, d'où son nom de modèle autorégressif tout-pôle. Cette fonction de

transfert est exprimée sous la forme d'un polynôme, appelé polynôme de prédiction. Les coefficients de ce polynôme modélisent le conduit vocal excité par un signal inconnu et dont les caractéristiques sont supposées. L'enjeu de l'analyse par prédiction linéaire consiste à déterminer les coefficients à partir du signal de parole, de manière à obtenir une bonne estimation de ses propriétés temporelles et spectrales. Étant donné la nature non stationnaire du signal de parole, les coefficients de prédiction sont estimés sur une courte durée du signal. L'approche la plus courante consiste à choisir les coefficients qui minimisent l'énergie de prédiction.

6.2.3.1.3. Méthodes

Deux algorithmes permettent de minimiser l'erreur moyenne quadratique de prédiction, la méthode de covariance et la méthode d'autocorrélation.

Pour obtenir les coefficients de prédiction, la méthode de covariance cherche à minimiser l'erreur moyenne quadratique de prédiction sur un court segment de la forme d'onde. Cette méthode a été présentée par ATAL en 1971 et doit son nom à la similarité entre la matrice utilisée et la matrice de covariance [ATAL, 1971 IN : SCHAFER ET RABINER, 1975].

Deux méthodes ont été proposées pour le calcul de la fonction d'autocorrélation à court terme : la méthode par vraisemblance maximale, développée par ITAKURA ET SAITO en 1970 et la méthode par filtrage inverse mise au point par MARKEL en 1972. La différence principale entre les deux approches est que la méthode d'autocorrélation nécessite l'utilisation d'un fenêtrage explicite, contrairement à la méthode de covariance, ce qui a pour conséquence la difficulté de mesure précise de la largeur des formants avec la méthode d'autocorrélation [MARKEL, 1972 IN : SCHAFER ET RABINER, 1975].

6.2.3.1.4. Estimation des fréquences formantiques

Les coefficients de prédiction linéaire permettent principalement d'estimer les formants. L'estimation des fréquences formantiques passe par la détermination du nombre approprié de coefficients du polynôme de prédiction. Celui-ci dépend du nombre de formants recherché et de la fréquence d'échantillonnage du signal analysé. Une bonne approximation consiste à compter une paire de coefficients pour la modélisation de chaque formant. En général lors de l'application de cette règle, les fréquences formantiques correspondent aux racines du polynôme de prédiction qui peuvent être obtenues par factorisation. Comme dans toute analyse formantique, la difficulté consiste à attribuer à un coefficient le rang correct du formant. Plusieurs algorithmes permettent néanmoins d'y parvenir [SCHAFER ET RABINER, 1975].

6.2.3.1.5. Estimation de la fréquence fondamentale

Cette information peut être obtenue par la méthode d'autocorrélation par filtrage inverse. Elle se fonde sur le filtrage inverse du signal et analyse la périodicité de la source estimée. Les coefficients de prédiction du filtre de transfert sont obtenus à partir du signal filtré dans la bande de 0 à 900 Hz. Le traitement par un filtre inverse du filtre de transfert permet d'obtenir une estimation de la source glottique. Finalement l'amplitude la plus élevée, et supérieure à un seuil

donné, est recherchée dans le résultat du calcul de la fonction d'autocorrélation sur l'estimation de la source glottique.

6.2.3.1.6. Application à la reconnaissance de locuteurs

ATAL a étudié en détail l'utilisation de la prédiction linéaire pour la reconnaissance de locuteurs. Il a comparé plusieurs paramétrisations dérivées de la prédiction linéaire comme les coefficients de réflexion, les coefficients d'autocorrélation, les coefficients de fonction d'aire et les coefficients cepstraux. Il a montré la supériorité des coefficients cepstraux, par rapport aux autres représentations du signal de parole, en obtenant un taux d'identification de 98% avec des échantillons de test de plus de 0,5 s, sur une base de données de dix locutrices enregistrées en deux sessions distantes de 27 jours [ATAL, 1974]. FURUI ET ITAKURA sont parvenus à des résultats similaires avec une base de données de neuf locuteurs, enregistrés sur une période de trois mois, en utilisant des coefficients de corrélation partielle (PARCOR) dérivés d'une analyse par prédiction linéaire [FURUI ET ITAKURA, 1973]. SAMBUR s'est intéressé à la prédiction linéaire orthogonale, suite à l'observation expérimentale de la redondance des coefficients de prédiction linéaire. Cette redondance implique qu'une analyse conventionnelle en composante principale peut être appliquée pour réduire l'espace dimensionnel de la prédiction linéaire. Deux paramétrisations dérivées ont été comparées aux coefficients de prédiction linéaire orthogonaux, les coefficients *partial correlation* orthogonaux (PARCOR) et les coefficients de fonction d'aire orthogonaux. L'utilisation des deux paramétrisations dérivées a apporté les meilleurs résultats : sur une base de données de 21 locuteurs, un taux d'identification de 99% a été obtenu en mode dépendant du texte et un taux de 94% en mode indépendant du texte. Lorsque les échantillons de parole transitent par une ligne téléphonique locale, ce taux diminue à 87% d'identification correcte [SAMBUR, 1979].

6.2.3.2. Analyse homomorphique

6.2.3.2.1. Principe

L'intérêt principal de l'analyse homomorphique réside dans sa capacité à séparer la contribution de la source de celle du conduit vocal par une opération de déconvolution. L'hypothèse de l'absence de couplage entre la source glottique et le conduit vocal facilite grandement le traitement du modèle, bien que ce postulat ne soit qu'approximativement vérifié [OPPENHEIM, 1968 IN : THEVENAZ, 1993]. Le domaine de définition du cepstre est un axe temporel gradué en unités de quérrence. C'est une description temporelle, définie comme l'inverse des fréquences que l'on trouve dans le signal.

Le cepstre complexe d'un signal s'obtient par le calcul de la transformée en z inverse du logarithme du spectre du signal, le cepstre réel étant la transformée en z inverse du logarithme du spectre d'amplitude du signal [THEVENAZ, 1990]. En pratique trois méthodes ont été développées pour calculer le cepstre réel du signal et estimer la période de la fréquence fondamentale et des fréquences formantiques.

6.2.3.2.2. Méthodes

La plus classique est fondée sur la transformée de Fourier à court terme. Le spectre de puissance à court terme du signal est calculé par transformée de Fourier discrète et exprimé en valeur logarithmique. Le résultat de cette transformation est ramené dans le domaine temporel par transformée de FOURIER discrète inverse, pour trouver les coefficients cepstraux [RABINER ET SCHAFER, 1978].

La deuxième méthode diffère de la précédente par le moyen de calculer le spectre de puissance à court terme. Il est assuré par un banc de filtres, avec calcul de la puissance pour chaque canal, plutôt que par transformée de Fourier discrète. Ces deux approches demeurent extrêmement proches car la transformée de Fourier à court terme peut être formulée sous la forme d'un banc de filtres uniforme, avec décimation des signaux de sous-bandes, et la possibilité de grouper les bandes de la transformée de Fourier permet de simuler un banc de filtres à largeur de bandes non uniforme [RABINER ET JUANG, 1993]. Le groupement de bandes de la transformée de Fourier à court terme ne correspond pas à un véritable banc de filtres, mais permet simplement d'obtenir une estimation de la puissance d'un banc de filtres équivalent [CAPPE, 1995].

La dernière approche utilise un modèle autorégressif du signal, qui fournit une première description paramétrique du contenu spectral du signal. Une formule de conversion permet ensuite de calculer les coefficients cepstraux à partir du modèle autorégressif [RABINER ET SCHAFER, 1978].

6.2.3.2.3. Application

Tous les coefficients cepstraux ne présentent pas une variabilité intralocuteur de même ampleur ; celle-ci décroît avec l'ordre des coefficients. Dans le domaine de la reconnaissance de locuteurs, l'influence des coefficients qui présentent la plus forte variabilité intralocuteur est réduite par une distance de Mahalanobis, correspondant à une distance euclidienne pondérée du fait de la décorrélation des coefficients cepstraux [SOONG ET ROSENBERG, 1988]. SOONG a aussi montré qu'il est pertinent de procéder au calcul des coefficients sur toutes les fenêtres où le signal est présent, voisé ou non [SOONG ET AL., 1987]. Par contre il est utile d'exclure les fenêtres où le signal de parole est absent, surtout lorsque le canal de transmission est de mauvaise qualité et susceptible de varier [NAIK ET AL., 1989 ; GISH, 1990 ; REYNOLDS, 1994].

6.2.3.3. Recherche de paramètres dérivés plus robustes

6.2.3.3.1. Intégration des découvertes psycho-acoustiques

Les effets conjugués des différences de l'environnement acoustique, du microphone et du canal de transmission téléphonique, et, dans le domaine criminalistique, du système d'enregistrement, affectent la qualité de l'extraction de caractéristiques pertinentes pour la reconnaissance de locuteurs et de parole. La découverte des principes psycho-acoustiques qui gouvernent l'audition humaine a conduit les chercheurs vers des stratégies d'extraction basées sur un certain mimétisme du système auditif humain.

La résolution spectrale de l'oreille humaine permet la différenciation d'environ 140 degrés de hauteur entre 0 et 500 Hz et d'environ 480 degrés de hauteur entre 500 Hz et 16 KHz, qui croissent proportionnellement à la fréquence. Les propriétés de cette sensibilité résultent de la structure de la membrane basilaire, large et flasque au sommet du limaçon et étroite et rigide à sa base, qui effectue une analyse fréquentielle mécanique tonotopique. Comme la relation entre la tonie et le lieu d'excitation principale de la membrane basilaire est linéaire, la résolution spectrale du système auditif humain décroît avec la fréquence ; elle est décrite par l'échelle psychoacoustique Mel, graduée de 0 à 2400 Mel. Dans le domaine de la perception artificielle, la meilleure approximation de ce mécanisme est donnée par un banc de filtres dont les bandes, appelées « bandes critiques », chevauchent et sont réparties selon la courbe de sensibilité tonotopique de l'oreille humaine.

La simulation du seuil de l'audition humaine, qui varie de 0 à 40 dB, peut quant à elle être réalisée par un filtre de correction de l'intensité. Pour une analyse de signaux sonores d'intensité modérée entre 0 et 5 KHz, MAKHOUL ET COSELL proposent par exemple une fonction de transfert avec des asymptotes de 12 dB par octave entre 0 et 400 Hz, de 0 dB par octave entre 400 et 1200 Hz, de 6 dB par octave entre 1200 et 3100 Hz et de 0 dB par octave entre 3100 Hz à la fréquence de Nyquist [MAKHOUL ET COSELL, 1976 IN : HERMANSKY, 1990]. Finalement, la relation psychoacoustique entre la pression acoustique d'un son et son intensité perçue, appelée loi de puissance de l'audition, n'est pas linéaire [STEVENS, 1957 IN : HERMANSKY, 1990]. Dans le domaine de la perception artificielle, l'application de cette loi passe par une approximation, sous forme de compression de l'amplitude du signal en calculant sa racine cubique.

6.2.3.3.2. Paramètres dérivés de la prédiction linéaire

Une des inconsistances du modèle autorégressif tout-pôle, utilisé pour l'analyse par prédiction linéaire, réside justement dans le fait qu'il approche de manière équivalente toutes les fréquences de la bande passante analysée. Plusieurs prétraitements du signal par des fonctions de distorsion spectrale ont été proposés, notamment par MAKHOUL ET COSELL ou STRUBE, qui proposent une distorsion spectrale selon l'échelle Mel [MAKHOUL ET COSELL, 1976 IN : HERMANSKY, 1990 ; STRUBE, 1980 IN : HERMANSKY, 1990].

HERMANSKY a étudié une classe de techniques de transformées spectrales qui modifient le spectre de puissance à court terme avant son approximation par le modèle autorégressif. Cette méthode, appelée prédiction linéaire perceptuelle, *perceptual linear prediction (PLP)*, consiste à procéder successivement à un filtrage en bandes critiques du spectre à court terme, à une correction de l'intensité et à la compression de l'amplitude du signal avant de procéder à l'analyse par prédiction linéaire (Figure VI.3.) [HERMANSKY, 1990].

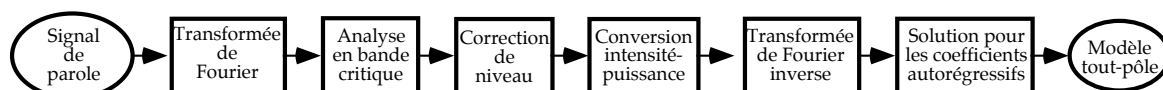


Figure VI.3. Prédiction linéaire perceptuelle [HERMANSKY, 1990]

Plusieurs autres méthodes robustes d'extraction des coefficients de prédiction linéaire, basées sur la minimisation de différentes fonctions objectives, ont été envisagées. Cependant les performances des différentes solutions proposées dépendent du type de dégradation du signal et aucune méthode aussi universelle que la prédiction linéaire perceptuelle n'a pu être dégagée [RAMACHANDRAN *ET AL.*, 1995].

6.2.3.3.3. Paramètres dérivés de l'analyse homomorphique

La simulation de la résolution spectrale de l'oreille humaine a aussi été adoptée pour l'analyse homomorphique, en calculant les coefficients cepstraux à partir d'un signal préalablement analysé dans un banc de filtres « en bandes critiques » [RABINER ET JUANG, 1993]. Les paramètres obtenus sont appelés coefficients cepstraux en échelle fréquentielle Mel ou *Mel Frequency Cepstrum Coefficients* (MFCC) (Figure VI.4.).

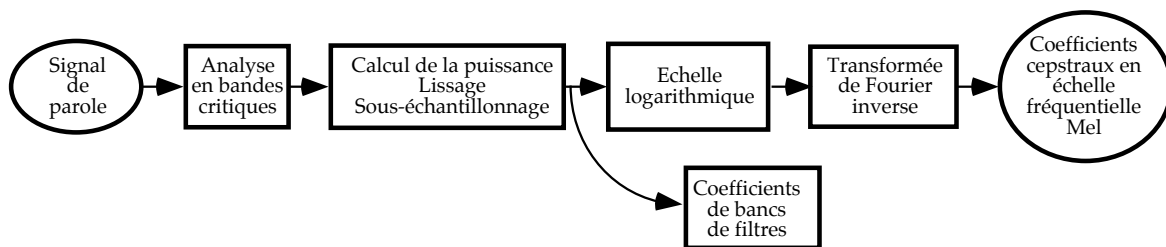


Figure VI.4. Coefficients cepstraux en échelle fréquentielle Mel [RABINER ET JUANG, 1993]

Une autre méthode consiste à analyser la variation des paramètres cepstraux dans des fenêtres proches de la fenêtre à court terme, pour extraire du signal des paramètres dynamiques appelés Δ -cepstraux [FURUI, 1981A ; SOONG ET ROSENBERG, 1988]. Leur efficacité est nettement inférieure aux paramètres cepstraux lorsqu'ils sont utilisés seuls, mais conduisent à une substantielle amélioration des performances lorsqu'ils sont combinés aux paramètres cepstraux instantanés [TSENG *ET AL.*, 1992]. L'intérêt principal des paramètres Δ -cepstraux réside dans leur insensibilité aux variations linéaires du canal de transmission [SOONG ET ROSENBERG, 1988]. Cependant l'effet des paramètres variationnels n'est significatif que dans des applications dépendantes du texte [CAPPE, 1995].

6.2.3.4. Techniques de normalisation

Une seconde stratégie consiste à atténuer les effets parasites par une normalisation du signal, afin d'extraire des caractéristiques plus proches de celles présentes dans un signal de parole non dégradé [MAMONNE *ET AL.*, 1996]. Les techniques de suppression de bruit par normalisation spectrale améliorent souvent l'intelligibilité, mais rarement les performances de reconnaissance [ASSALEH ET MAMONNE, 1994]. La normalisation spectrale par soustraction du spectre moyen à long terme fait exception à cette règle [BOLL, 1979 *IN* : MAMONNE *ET AL.*, 1996 ; ROSENBERG ET SOONG, 1991 ; FURUI, 1994].

Dans le domaine cepstral une opération de filtrage linéaire se traduit par une modification additive, par conséquent la compensation de l'effet de ce filtrage est théoriquement plus simple que dans le domaine spectral [CAPPE, 1995]. La soustraction de la moyenne cepstrale à long terme

contribue à une amélioration des performances lorsque les caractéristiques du canal de transmission des signaux comparés sont différents, mais diminue ces performances lorsque les caractéristiques du canal sont identiques pour les deux signaux comparés [MAMONNE *ET AL.*, 1996]. La soustraction de la moyenne cepstrale à court terme, susceptible de supprimer non seulement l'effet du canal de transmission, mais aussi de l'information dépendante du locuteur, ne conduit pas à une diminution des performances, mais se révèle une technique de normalisation efficace [PAWLEWSKI ET DOWNEY, 1996].

La technique spectrale relative (RASTA) tire parti du fait que l'évolution temporelle de la partie du signal liée aux composantes non linguistiques ne correspond pas à l'évolution temporelle de la partie liée aux composantes linguistiques, en supprimant les composantes spectrales dont l'évolution est plus lente ou plus rapide que celle du tractus vocal. Contrairement à la soustraction de la moyenne cepstrale à long terme, qui supprime la composante continue du logarithme du spectre à court terme, cette technique influence le spectre du signal de parole de manière plus complexe en accentuant les transitions spectrales [HERMAN SKY ET MORGAN, 1994].

La technique RASTA peut être combinée à la méthode de prédiction linéaire perceptuelle (PLP) pour calculer la fonction de transfert du filtre. Dans ce cas, les trajectoires temporelles des composantes spectrales sont filtrées pour supprimer les composantes non linguistiques du spectre. Ce spectre filtré est ensuite approché par un modèle autorégressif. Cette technique peut être appliquée directement aux coefficients cepstraux issus de la fonction de transfert de prédiction linéaire calculée par la méthode conventionnelle d'autocorrélation [HERMAN SKY ET MORGAN, 1994].

Toutes les études focalisées sur la comparaison de caractéristiques dépendantes du locuteur montrent qu'actuellement les coefficients cepstraux en échelle Mel (MFCC) et les coefficients issus de la prédiction linéaire perceptuelle (PLP) alliés à leurs extensions RASTA et leurs dérivées de premier ordre sont considérés comme les paramètres les plus robustes pour la reconnaissance de locuteurs indépendante du texte en milieu bruyé [OPENSHAW *ET AL.*, 1993 ; KAO *ET AL.*, 1993 ; REYNOLDS, 1996 ; VAN VUUREN, 1996 ; FURUI, 1997].

6.2.4. Conclusion

En 1994, FURUI reprend presque la conclusion de ROSENBERG et SOONG de 1991, en constatant que la recherche se focalise actuellement plus sur l'amélioration des mesures de similarité que sur la recherche de méthodes d'analyse du signal de parole plus efficaces :

«Les progrès récents obtenus dans le domaine de la reconnaissance de locuteurs sont principalement dus à l'amélioration des techniques utilisées pour modéliser et décrire les caractéristiques mesurées pour chaque locuteur. Ces progrès n'ont pas forcément permis d'accroître ou d'améliorer nos connaissances en ce qui concerne les particularités propres à chaque locuteur et la manière de les extraire du signal de parole » [ROSENBERG ET SOONG, 1991 ; FURUI, 1994].

6.3. Mesure de similarité

Le principe qui sous-tend toutes les méthodes de mesure de similarité et de probabilité est la modélisation des caractéristiques dépendantes du locuteur extraites lors de l'analyse du signal. Avant le développement des algorithmes probabilistes, la mesure de similarité était ramenée à un problème de reconnaissance de formes, où la distance était calculée sur la base de formes extraites des données de test et comparées aux formes de référence, calculées pour chaque locuteur à partir des données d'entraînement. Les méthodes actuelles s'attachent à décrire la distribution statistique des caractéristiques extraites du signal et peuvent être définies comme des méthodes de modélisation probabiliste de données multidimensionnelles.

De chaque fenêtre d'analyse sont extraits un certain nombre de paramètres qui peuvent être considérés comme une quantité vectorielle. Par exemple, l'extraction de douze paramètres conduit à la définition d'un vecteur dans un espace à douze dimensions et chaque paramètre constitue une coordonnée du vecteur mesuré. Cette interprétation est particulièrement justifiée pour les coefficients de prédiction linéaire et les coefficients cepstraux, qui sont des quantités homogènes.

Les techniques de reconnaissance de formes se répartissent en deux catégories : les méthodes paramétriques, dans lesquelles la forme de la distribution de la quantité vectorielle est supposée connue et les méthodes non paramétriques, qui ne s'appuient sur aucun modèle connu de forme de distribution. A ceci s'ajoute une distinction importante dans le domaine de la reconnaissance de locuteurs, entre les méthodes séquentielles, qui tiennent compte de l'ordre de mesure des vecteurs d'observation, applicables en mode dépendant du texte, et les méthodes globales, où cet ordre n'est pas considéré comme significatif, applicables en mode dépendant et indépendant du texte [GISH ET SCHMIDT, 1994 ; CAPPE, 1995].

6.3.1. Approches primaires

6.3.1.1. Discrimination par la valeur moyenne

6.3.1.1.1. Principe

Cette technique non paramétrique globale consiste à caractériser la distribution des caractéristiques vectorielles mesurées par leur valeur moyenne. Comme les caractéristiques sont principalement liées au spectre à court terme, ce type d'analyse est souvent désigné par le terme de « spectre moyen à long terme » [CAPPE, 1995]. En mode de reconnaissance de locuteurs indépendante du texte, la durée des échantillons de parole devrait idéalement atteindre de plusieurs secondes à quelques minutes, de manière à modéliser la voix et non des artefacts locaux. L'efficacité de la méthode dépend directement des vecteurs de caractéristiques et de la mesure de distance choisie pour effectuer la comparaison de ces vecteurs [GISH ET SCHMIDT, 1994].

La métrique la plus simple utilisée pour la réalisation d'un classificateur de distance minimale est la mesure de la distance euclidienne. La mesure de distance peut également être basée sur la mesure de corrélation entre les deux vecteurs de caractéristiques comparés. Il est

possible de modifier ces deux mesures de distance en pondérant chaque dimension de l'espace vectoriel par la valeur de l'inverse de sa variance entre différents énoncés du même locuteur, de manière à privilégier l'influence des composantes les plus fiables du vecteur de caractéristiques [BUNGE, 1979].

Le rendement optimal d'un classificateur de distance minimale est obtenu lorsque les vecteurs de base de l'espace vectoriel ne sont pas corrélés, ce qui correspond à une situation particulière rarement atteinte lors du choix des paramètres. Le calcul d'une matrice de corrélation des vecteurs de base permet de connaître leur taux de corrélation et de modifier leur position par rotation dans l'espace vectoriel jusqu'à obtenir leur orthogonalité par une analyse en composantes principales ou une transformation de Karhunen-Loeve [CORSI, 1982 ; O'SHAUGNESSY, 1986]. L'application du classificateur de distance minimale à ces données orthogonales, en considérant la distance euclidienne, correspond à la classification de Mahalanobis [BRICKER *ET AL.*, 1971 ; BUNGE, 1979].

Notamment à cause de la complexité mathématique de l'analyse en composantes principales et de la transformation de Karhunen-Loeve, d'autres métriques ont été étudiées, comme le classificateur de risque minimal, exploitant le théorème de Bayes, ou le classificateur d'Anderson et Bahadur, exploitant différentes matrices de covariance [BUNGE, 1979, CORSI, 1982 ; KRZYSZKO *ET AL.*, 1973 ; KACZMAREK ET KRZYSZKO, 1973 ; CALINSKI ET KACZMAREK, 1968].

6.3.1.1.2. Application

Le calcul de la moyenne du spectre à court terme sur l'ensemble du signal analysé permet d'obtenir deux caractéristiques exploitables en mode indépendant du texte, le spectre moyen à long terme ainsi que son écart type. Une étude de FURUI *ET AL.* montre que l'écart type des données spectrales à long terme est pratiquement double de celui des données spectrales à court terme [FURUI *ET AL.*, 1972]. Dans une expérience menée conjointement aux États-Unis et en Pologne, HOLLIEN ET MAJEWSKI concluent qu'une telle approche peut être envisagée en cas d'élocution normale, voire sous stress, mais pas en cas de voix déguisée [HOLLIEN ET MAJEWSKI, 1977]. D'autre part, GUBRYNOWICZ met en évidence le manque de robustesse de la méthode aux variations du canal de transmission, notamment en cas de variation de la bande passante entre les échantillons comparés [GUBRYNOWICZ, 1973].

Plusieurs techniques de normalisation ont été suggérées en vue de compenser les variations du canal de transmission, notamment les effets induits par l'utilisation de lignes téléphoniques différentes. GLENN ET KLEINER, tout comme DODDINGTON, normalisent le vecteur des données spectrales obtenues en sortie du banc de filtres par la somme des sorties du filtre pour chaque mesure [GLENN ET KLEINER, 1968 *IN* : ROSENBERG, 1976B, DODDINGTON 1974 *IN* : ROSENBERG, 1976B]. Cette normalisation a pour effet de stabiliser les mesures par rapport aux variations du niveau du signal. Une technique de filtrage inverse connue sous le nom de distance d'Itakura a aussi été proposée ; le spectre à long terme sert alors à la définition d'un filtre inverse du second ordre qui caractérise la distribution spectrale grossière du signal d'entrée [TOHKURA, 1986].

Le spectre moyen à long terme reste toutefois une réduction extrême des caractéristiques spectrales des énoncés d'un locuteur et le pouvoir discriminatoire de certaines séquences du spectre à court terme utilisé en mode dépendant du texte lui échappe [FURUI, 1997]. Le même constat d'insuffisance peut être tiré à propos des caractéristiques temporelles à long terme [MAJEWSKI *ET AL.*, 1979]. La meilleure alternative consiste à utiliser le spectre moyen à long terme comme élément de normalisation plutôt que comme caractéristique dépendante du locuteur, puisqu'il se révèle être trop sensible aux variations des caractéristiques dépendantes du locuteur et du canal de transmission [ROSENBERG ET SOONG, 1991, FURUI, 1994].

6.3.1.2. Alignement temporel par programmation dynamique

6.3.1.2.1. Principe

L'alignement temporel par programmation dynamique, *Dynamic Time Warping* (DTW), est une méthode non paramétrique séquentielle, applicable en mode dépendant du texte [CAPPE, 1995]. Le principe a été décrit par Flanagan et implémenté originellement par DODDINGTON en 1970, sous le nom de *warping function* [DODDINGTON, 1970]. Chaque énoncé est représenté par une séquence de vecteurs caractéristiques généralement liés au spectre à court terme. La variation temporelle de l'énoncé de référence et de l'énoncé de test est normalisée par alignement non linéaire de la séquence des vecteurs caractéristiques en utilisant l'algorithme de programmation DTW [FURUI, 1997]. La distance euclidienne cumulée, calculée entre l'échantillon de référence et l'échantillon de test, sert à la classification [DODDINGTON, 1985].

6.3.1.2.2. Application

Suite aux travaux de DODDINGTON, l'alignement temporel par programmation dynamique a été étudié par LUMMIS, qui a utilisé la valeur absolue de la différence entre les énergies des énoncés comparés pour procéder à leur alignement [LUMMIS, 1973]. En 1976, ROSENBERG décrit un système de reconnaissance développé chez *Bell Telephone Laboratories*[®], exploitant l'algorithme DTW pour aligner le contour de F_0 . Une base de données de plus de 100 locuteurs enregistrés par téléphone a été utilisée pour le test. La référence de chacun des locuteurs est obtenue par l'enregistrement de cinq énoncés en une seule session. Cinquante enregistrements de test de chaque locuteur ont été recueillis sur une période de cinq mois et les performances reportées indiquent un taux d'égale erreur ⁷⁵ d'environ 10% pour la tâche d'identification [ROSENBERG, 1976B ; BIMBOT, 1993].

En 1981, Furui présente un système de reconnaissance où, pour chacun des 50 locuteurs de la base de données, la référence est constituée d'une ou de plusieurs répétitions de chaque mot d'un vocabulaire déterminé. Les énoncés de test sont comparés avec l'algorithme DTW à la concaténation des mots de référence correspondants. Cette méthode permet d'obtenir un taux d'erreur d'un peu plus de 2% en tâche de vérification.

⁷⁵ *supra* : 3.5.3. Quantification des taux d'erreur de type I et de type II

En 1985, la version la plus récente du système de contrôle d'accès du centre de calcul de *Texas Instruments*[®], également basée sur l'algorithme DTW, aboutissait à des taux d'erreur de type I et de type II inférieurs à 1% [DODDINGTON, 1985].

Malgré une amélioration constante de la technique jusqu'à atteindre, dans l'étude de BERNASCONI, un taux de vérification supérieur à 99,9% sur une population de 22 locuteurs, l'alignement temporel par programmation dynamique a peu à peu été abandonné au profit de modèles séquentiels statistiques comme les modèles de Markov cachés, moins rigides et plus robustes vis-à-vis de la variabilité inhérente au signal de parole [BERNASCONI, 1990 ; CAPPE, 1995].

6.3.2. Approches actuelles

6.3.2.1. Classification gaussienne

La classification gaussienne est une technique paramétrique globale basée sur l'hypothèse que les paramètres mesurés suivent une répartition gaussienne. Cette distribution gaussienne est multidimensionnelle puisque les paramètres sont définis dans un espace vectoriel multidimensionnel. Les paramètres estimés sont, en plus du vecteur moyen, la matrice de covariance des paramètres [CAPPE, 1995].

Cette modélisation gaussienne permet de calculer un rapport de vraisemblance entre l'énoncé de test et l'énoncé de référence à partir de la forme analytique des distributions gaussiennes représentant les paramètres, lorsque l'hypothèse que les paramètres sont statistiquement indépendants est admise. Le calcul de la vraisemblance fait intervenir à la fois la moyenne et la matrice de covariance des paramètres, mais dans le cadre d'une application où les enregistrements sont réalisés par le téléphone, le terme lié à la moyenne des mesures est peu significatif [GISH ET AL., 1985 ; GISH ET AL., 1986].

Devant les difficultés posées par une modélisation de l'effet du canal de transmission en l'absence d'information le concernant, une modification préconisée du classificateur gaussien consiste à ne considérer que la partie de la vraisemblance qui dépend de la matrice de covariance des données [KRASHNER ET AL., 1984 ; GISH ET AL., 1985 ; GISH ET AL., 1986 ; GISH 1990]. Cette modification du classificateur gaussien conduit à une approche visant à mesurer la similarité existant entre les matrices de covariance [CAPPE, 1995]. Diverses variantes de cette approche, comme les mesures statistiques du second ordre, ont été présentées ; toutes font, au moins implicitement, référence au modèle gaussien multidimensionnel, dans le sens où elles supposent que la matrice de covariance permet de rendre efficacement compte de la répartition des données [GISH, 1990 ; BIMBOT, 1993 ; BIMBOT ET MATHAN, 1994 ; GISH ET SCHMIDT, 1994 ; BIMBOT ET AL., 1995]. Il faut cependant remarquer que l'hypothèse gaussienne n'est pas vérifiée en pratique pour des paramètres tels que les coefficients cepstraux ou les coefficients de prédiction linéaire.

L'attrait pour les techniques inspirées du modèle gaussien tient au fait qu'elles sont peu coûteuses en temps de calcul, nécessitent l'estimation d'un petit nombre de paramètres et peuvent déjà être mises en œuvre avec une durée d'apprentissage d'une quinzaine de secondes de parole.

La comparaison d'énoncés de référence et d'énoncés de test contemporains par des mesures statistiques du second ordre fournissent d'excellentes performances d'identification de locuteurs en mode indépendant du texte, sur la base d'enregistrements de haute fidélité, d'enregistrements dont la bande passante est limitée artificiellement à 4 KHz, mais pas sur la base d'enregistrements téléphoniques [BIMBOT, 1993 ; BIMBOT ET MATHAN, 1994].

Le classificateur gaussien est un cas particulier du mélange de fonctions de densité gaussiennes⁷⁶ et les résultats présentés par plusieurs chercheurs montrent que les performances du classificateur gaussien sont systématiquement surpassées par les performances du modèle par mélange de fonctions de densité gaussiennes. Ce constat démontre la meilleure capacité de ce modèle à représenter la distribution réelle des paramètres dérivés des analyses par prédiction linéaire et homomorphique [ROSE ET REYNOLDS, 1990 ; MATSUI ET FURUI, 1992 ; TSENG ET AL., 1992].

6.3.2.2. Représentation par quantification vectorielle

La quantification vectorielle (*Vector Quantization*, VQ) est une méthode non paramétrique globale, applicable en mode dépendant et indépendant du texte, qui permet de décrire un ensemble de données par un faible nombre de vecteurs formant un dictionnaire (*codebook*) associé aux données. Le dictionnaire de quantification des spectres d'un locuteur est calculé de manière à ce que la distance entre un vecteur issu des données et son plus proche voisin dans le dictionnaire soit la plus faible possible, en d'autres termes que la quantification vectorielle crée le moins de distorsions sur la parole de ce locuteur [MAKHOUL ET AL., 1985]. La quantification vectorielle est une technique de groupage (*clustering*) d'autant plus adaptée que la parole présente naturellement des points d'accumulation autour desquels la densité de vecteurs issus des données est importante [CAPPE, 1995].

La quantification vectorielle est généralement réalisée par une méthode d'optimisation successive de dictionnaires de taille croissante appelée *binary (splitting) K-means*, qui permet de contourner le délicat problème de l'initialisation de l'algorithme de recherche itérative des vecteurs du dictionnaire [MAKHOUL ET AL., 1985]. Pour la reconnaissance de locuteurs, la mesure de similarité entre deux jeux de mesures consiste à évaluer la distorsion moyenne d'un des deux ensembles de mesures en utilisant le dictionnaire optimisé par quantification vectorielle pour l'autre jeu de mesures [CAPPE, 1995].

La caractérisation de la distribution des données obtenue par la quantification vectorielle est proche de celle fournie par un modèle de mélange de fonctions de densité gaussiennes et les performances de ces deux méthodes sont comparables. Lorsque les données disponibles pour l'apprentissage sont suffisantes, il semble toutefois que le modèle de mélange de fonctions de densité gaussiennes soit plus robuste. A l'opposé lorsque les enregistrements utilisés pour l'apprentissage durent moins de 20 s, la quantification vectorielle semble fournir une description

⁷⁶ *infra* : 6.3.2.3. Modélisation par mélange de fonctions de densité gaussiennes

plus fiable que le modèle de mélange gaussien, qui nécessite l'estimation d'un grand nombre de paramètres [MATSUI ET FURUI, 1992 ; CAPPE, 1995].

Le volume de données nécessaire à l'apprentissage peut être estimé suite à l'observation que les performances de reconnaissance ne s'améliorent que très peu lorsque le nombre de vecteurs du dictionnaire dépasse 2^6 ou 2^7 et l'algorithme de quantification vectorielle ne fonctionne de manière satisfaisante que si l'on dispose d'au moins 20 à 50 fois plus de vecteurs de données que de vecteurs du dictionnaire, ce qui correspond à environ 20 s de parole avec une extraction des paramètres tous les 10 ms [SOONG ET AL., 1985 ; ROSENBERG ET SOONG, 1986 ; MATSUI ET FURUI, 1992]. En pratique, une diminution très nette des performances est observée lorsque la durée des enregistrements d'apprentissage devient inférieure à une dizaine de secondes [MATSUI ET FURUI, 1991]. MATSUI et FURUI ont réalisé des mesures de performance d'identification en mode indépendant du texte. La base de données comprend 13 locutrices et 23 locuteurs ; pour chacune et chacun, quinze phrases de 4 s ont été collectées en trois occasions sur une période de six mois. Les modèles ont été calculés à partir de la concaténation de dix phrases, alors que les cinq phrases restantes ont été utilisées pour le test, à partir d'enregistrements de test formés de phrases de 4 s dans une base de données. Les résultats s'échelonnent entre 86.9% et 95.4% d'identification correcte, selon la taille du dictionnaire et la vitesse d'élocution des locuteurs. Les meilleurs résultats ont été obtenus pour une élocution à vitesse normale, avec le plus grand dictionnaire testé, constitué de 512 vecteurs [MATSUI ET FURUI, 1992].

6.3.2.3. Modélisation par mélanges de fonctions de densité gaussiennes

La modélisation par mélange de fonctions de densité gaussiennes, *Gaussian Mixture Models* (GMM), est une méthode paramétrique globale. Elle consiste à supposer que la distribution des caractéristiques dépendantes du locuteur peut être décrite par une fonction de densité de probabilité gaussienne multidimensionnelle, sous la forme d'un vecteur de moyennes et d'une matrice de covariance. Le vecteur de moyennes représente les valeurs attendues des caractéristiques analysées, alors que la matrice de covariance rend compte des corrélations et de la variabilité de ces caractéristiques [CAPPE, 1995 ; REYNOLDS, 1995A ; REYNOLDS, 1995B]. L'étude de la structure des fonctions de densité de probabilité gaussienne composant le mélange a souvent conduit les chercheurs à la simplifier en considérant qu'elles possèdent toutes une matrice de covariance diagonale [MATSUI ET FURUI, 1992 ; TSENG ET AL., 1992 ; REYNOLDS, 1994]. La justification de cette simplification réside d'une part dans la difficulté d'estimation complète des matrices de covariance et, d'autre part, dans la faible corrélation des caractéristiques cepstrales et de prédiction linéaire analysées à l'heure actuelle. Cette approximation contribue cependant à une légère dégradation des performances de reconnaissance [TSENG ET AL., 1992].

Cette méthode de modélisation, connue dans le domaine de la reconnaissance de formes, est fondée sur l'hypothèse que les caractéristiques dépendantes du locuteur appartiennent à un ensemble de classes différentes, avec une probabilité d'appartenance propre à chaque classe. Le modèle GMM considère le cas particulier dans lequel la distribution des données suit une loi gaussienne à l'intérieur de chaque classe. Ce choix tient essentiellement au fait que la loi gaussienne appartient à la famille des lois de distribution exponentielles, pour lesquelles le

problème de l'identification des composantes du mélange se trouve simplifié [REDNER ET WALKER, 1984 ; CAPPE, 1995]. Cette approche semble adaptée aux caractéristiques du signal de parole et se rapproche de la caractérisation fournie par la quantification vectorielle. Elle diffère de cette dernière par la description de la distribution des caractéristiques autour de certains points d'accumulation, alors que la quantification vectorielle se contente de les mettre en évidence.

L'estimation des classes du modèle est par nature très complexe [DUDA ET HART, 1973]. Elle est réalisée par l'algorithme *Expectation-Maximisation* (EM) qui, par un processus itératif de prévision et de maximisation non supervisé, recherche les classes du modèle qui permettent de maximiser la distribution des caractéristiques analysées. Cependant, l'algorithme EM est susceptible de fournir de multiples solutions et, de plus, avec une convergence lente simplifiée [REDNER ET WALKER, 1984]. Plusieurs méthodes d'initialisation de l'algorithme d'apprentissage ont donc été proposées, soit de manière simple par partitions arbitraires [MATSUI ET FURUI, 1992], soit de manière plus élaborée à l'aide d'une détermination initiale des paramètres par une procédure de quantification vectorielle [ROSE ET AL., 1991]. La mesure de similarité calculée par la méthode GMM est la probabilité conditionnelle des caractéristiques de l'enregistrement de test, sachant celles du modèle ⁷⁷ [REYNOLDS, 1995A ; REYNOLDS, 1995B ; MOON, 1996].

L'algorithme EM est utilisé pour la modélisation dans des domaines où des facteurs inconnus influencent les résultats, comme l'économétrie, la médecine clinique et la sociologie. Dans le domaine du traitement de signal, les premières applications concernent la reconstruction d'images tomographiques et l'entraînement des modèles de Markov cachés dans le domaine de la reconnaissance de parole, et plus récemment la reconnaissance de formes et l'entraînement des réseaux neuromimétiques, la suppression de bruit ou la spectroscopie. L'algorithme EM est aussi lié aux algorithmes utilisés dans la théorie de l'information, car l'étape de prévision produit un résultat semblable au calcul de l'entropie [MOON, 1996].

Comme le montrent les évaluations des méthodes de reconnaissance de locuteurs effectuées par le *National Institute of Standards* (NIST) américain en 1996, 1997 et 1998, la modélisation par mélange de fonctions de densité gaussiennes représente l'état de l'art pour la reconnaissance de locuteurs en mode indépendant du texte, lorsque la quantité de données nécessaires à la constitution du modèle est suffisante [PRZYBOCKI ET MARTIN, 1998]. Lorsque la durée des énoncés utilisés pour la constitution du modèle est inférieure à 20 s, la méthode GMM semble moins efficace que la quantification vectorielle, compte tenu du nombre important de paramètres qu'il est nécessaire d'estimer [CAPPE, 1995 ; FURUI, 1997]. MATSUI et FURUI ont réalisé des mesures de performance d'identification en mode indépendant du texte dans les mêmes conditions que celles utilisées pour la quantification vectorielle. Les résultats obtenus sont comparables à ceux obtenus pour la quantification vectorielle. Ils s'échelonnent entre 85,8% et 95,6% d'identification correcte, selon la vitesse d'élocution et le nombre de gaussiennes utilisées pour le mélange. Les performances maximales sont obtenues pour la vitesse d'élocution normale, avec un mélange de 64 gaussiennes [MATSUI ET FURUI, 1992].

⁷⁷ *infra* : 7.2.3.4. Comparaison

6.3.2.4. Modélisation par modèles de Markov cachés

La modélisation par modèles de Markov cachés (*Hidden Markov Models*, HMM) est une approche paramétrique et séquentielle, puisqu'elle prend en compte certains aspects séquentiels du signal de parole ; elle s'est avérée très efficace notamment dans le cadre de la reconnaissance de la parole.

Le modèle de Markov caché est un modèle statistique séquentiel qui suppose que les caractéristiques observées forment une succession d'états distincts. Il est caractérisé par trois éléments : les probabilités initiales de se trouver dans chaque état, les probabilités de transition, qui décrivent les passages possibles entre les différents états, et les probabilités de sortie, qui représentent les distributions conditionnelles des caractéristiques observées en fonction de l'état du modèle. Les règles de transition permises entre les états définissent les différents types de modèles markoviens et le choix de la topologie du modèle dépend du mode de reconnaissance de locuteurs, dépendant ou indépendant du texte.

Le mode dépendant du texte fait appel à des modèles de topologie gauche-droite où les états correspondent aux mots du texte (Figure VI.5.) [ROSENBERG ET SOONG, 1991].

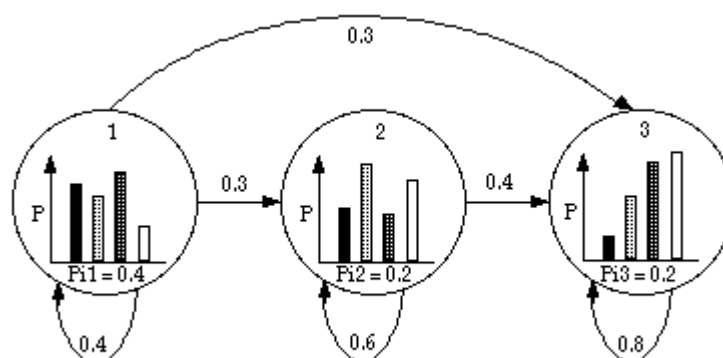


Figure VI.5. Modèle de Markov gauche-droite à 3 états : (P_i = probabilité d'état initial)

Le mode indépendant du texte requiert des modèles ergodiques ou des modèles gauche-droite où les états correspondent à des unités phonétiques connectées (Figure VI.6.) [SAVIC ET GUPTA, 1990].

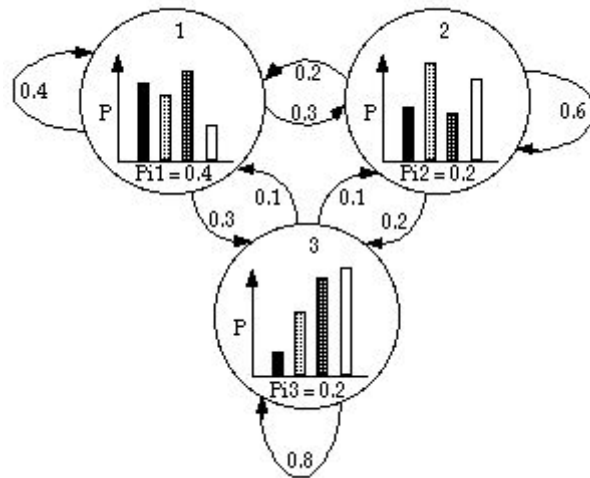


Figure VI.6. Modèle de Markov ergodique à 3 états (P_i = probabilité d'état initial)

Les aspects séquentiels de la parole pris en compte dans la modélisation par modèles de Markov cachés sont la source des excellentes performances de cette méthode en mode dépendant du texte lorsque le vocabulaire est fixe ou très contraint. En mode indépendant du texte, par contre, l'information apportée par les transitions entre états n'améliore pas les performances de reconnaissance de locuteurs et, dans ces conditions, les méthodes de reconnaissance basées sur des modèles de Markov ne concurrencent pas les performances des méthodes basées sur GMM, même si elles les approchent parfois [DE VETH ET BOURLARD, 1995 ; LAMEL ET GAUVAIN, 1998].

Dès lors la complexité du modèle de Markov ne se justifie pas, face à l'équivalent non séquentiel que représente le modèle par mélange de fonctions de densité gaussiennes ; cette modélisation est aussi définie comme un modèle de Markov continu où la distribution conditionnelle dans chaque état est un mélange de fonctions de densité gaussiennes. Les mesures de performance d'identification en mode indépendant du texte obtenues dans les mêmes conditions que celles utilisées pour la quantification vectorielle et la modélisation par mélange de fonctions de densité gaussiennes illustrent cette réalité. En effet les résultats obtenus sont nettement inférieurs à ceux obtenus avec les méthodes basées sur VQ et GMM. Ils s'échelonnent entre 74,7% et 88,3% d'identification correcte, selon la vitesse d'élocution, le nombre d'états et la taille du dictionnaire. Les performances maximales sont obtenues pour la vitesse d'élocution normale, avec un modèle à 4 états et un dictionnaire de 256 vecteurs [MATSUI ET FURUI, 1992].

6.3.2.5. Autres méthodes

Des méthodes représentant des généralisations séquentielles des méthodes globales que sont la quantification vectorielle et le classificateur gaussien ont été proposées. A l'image du fonctionnement de la modélisation par modèles de Markov cachés par rapport à la modélisation par mélange de fonctions de densité gaussiennes, la quantification matricielle et la modélisation autorégressive multidimensionnelle sont susceptibles d'apporter des améliorations dans le cadre des applications en mode dépendant du texte ou à vocabulaire restreint [ROSENBERG ET SOONG, 1991 ; CHEN ET AL., 1993 ; BIMBOT, 1993 ; FURUI, 1994].

L'application de réseaux de neurones artificiels, *artificial neural networks* (ANN's), pour la reconnaissance de locuteurs a été développée surtout au début des années 1990. Comme le montre notamment HENNEBERT, les classificateurs de type *Multi Layer Perceptron* (MLP) sont puissants et fournissent des réponses en termes de probabilité ; par contre cette méthode est non linéaire et l'explicitation du processus d'apprentissage et de classification du réseau neuronal demeure difficile [HENNEBERT, 1999]. Le phénomène de « boîte noire », lié à ce manque d'explicitation, rend délicate l'utilisation des réseaux de neurones artificiels en sciences forensiques pour l'instant.

6.4. Systèmes automatiques développés en sciences forensiques

6.4.1. Semi-Automatic Speaker Identification System (SASIS) - USA (1971 - 1975)

La seconde moitié du fond de 300'000 dollars alloué en 1971 par le *Law Enforcement Administration Assistance of the United States Department of Justice* (LEAA) pour la recherche sur la reconnaissance de locuteurs est dévolu au *Sensory Sciences Research Center of the Stanford Research Institute* (SRI) pour le développement d'un système semi-automatique de reconnaissance de locuteurs, la première moitié ayant été attribuée au *Department of Michigan State Police*, afin de procéder à la vérification des hypothèses de KERSTA ⁷⁸ [BECKER ET AL., 1973].

Le prototype mis au point au SRI se compose d'un premier étage de numérisation, suivi d'un étage de visualisation et d'édition manuelle du signal de parole, d'un étage d'extraction des caractéristiques et finalement d'un étage de mesure de similarité.

L'étage de numérisation est capable de traiter des échantillons de parole d'une durée allant jusqu'à six secondes, échantillonnés à 10 KHz et quantifiés sur 11 bits. Le deuxième étage du prototype est très évolué du point de vue ergonomique pour l'époque, puisqu'il permet la visualisation de la forme d'onde du signal de parole sur un écran et sa segmentation, rendue possible par l'utilisation d'une souris informatique. La sélection manuelle des six voyelles, éléments courts, mais discriminants, tend à rendre la méthode indépendante du texte et vise à diminuer l'intravariabilité par rapport à l'intervariabilité ; l'influence du phénomène de coarticulation est sous-estimée par les auteurs [BECKER ET AL., 1973].

Les méthodes d'extraction de caractéristiques intégrées dans le troisième étage sont basées sur l'analyse du spectre à court terme à partir de la transformée de Fourier rapide et sur la prédiction linéaire, profitant ainsi des travaux de ATAL et de MARKEL [ATAL, 1971 IN : BECKER ET AL., 1973 ; MARKEL, 1972 IN : BECKER ET AL., 1973].

La mesure de similarité est réalisée à l'aide de plusieurs métriques : la mesure de rapports de vraisemblance entre les caractéristiques en supposant leur indépendance, la mesure de la distance

⁷⁸ *supra* : 5.2.2. Tentative de validation de la méthode de KERSTA : l'étude de TOSI

euclidienne entre les caractéristiques, soit uniformément pondérées dans un espace multidimensionnel, soit pondérées par leur écart type lorsque celui-ci est très variable d'une caractéristique à l'autre, soit pondérées par le discriminant de Fischer, afin d'augmenter l'influence des caractéristiques les plus discriminantes.

La décision est considérée sous l'angle de la discrimination et le seuil de discrimination est défini partiellement par la machine et partiellement par l'opérateur, qui peut de cette manière prendre l'initiative d'une non-décision.

Les tests ont été effectués par des étudiants sans connaissance particulière en science de la parole, sur la base de 200 phrases prononcées par 100 locuteurs. Les résultats montrent que si les opérateurs définissent des seuils de manière à ne rendre aucune décision dans environ 30% des cas, les taux d'erreur de type I et de type II sont inférieurs à 1% en cas de décision.

Sur la base de ces résultats et d'une évaluation parallèle effectuée par *Texas Instruments*[®], qui a notamment mis en évidence le manque de robustesse du système lorsque les locuteurs sont enrhumés, les auteurs concluent que cette méthode est utilisable à des fins d'investigation par des opérateurs sans formation particulière. Par contre ils recommandent des études complémentaires sur des paramètres comme la coarticulation, la variabilité intersession et la langue parlée avant de proposer l'utilisation de cette méthode devant un tribunal [BECKER ET AL., 1973].

Ces conclusions ont conduit le LEAA à poursuivre les recherches en mandatant la firme *Rockwell International*[®], dont la division de recherche en électronique a développé le programme *Semi-Automatic Speaker Identification System (SASIS)*, visant à améliorer le système mis au point par le SRI. Ce programme s'articule sur plusieurs axes de recherche : l'enregistrement d'une base de données de 250 locuteurs, l'analyse des effets de la coarticulation et de la variation du canal de transmission sur la reconnaissance, la sélection des caractéristiques pertinentes et des métriques optimales pour la mesure de similarité [PAUL ET AL., 1975]. Après son développement, le projet a cependant été abandonné, pour des raisons de difficulté d'utilisation par des opérateurs non spécialistes des sciences de la parole et par manque de résultat dans des conditions forensiques réelles [KÜNZEL, 1994A ; POZA, 1999]

6.4.2. Automatic Recognition Of Speakers (AUROS) – Allemagne (1977)

Le projet AUROS, financé par le ministère allemand de la Recherche, a été développé par Ernst BUNGE, d'abord au laboratoire de recherche de Philips à Hambourg, puis au BKA.

Pour éviter les difficultés de la segmentation phonétique-acoustique du signal de parole et obtenir un fonctionnement de la méthode en mode indépendant du texte, l'analyse du signal de parole repose sur l'extraction de caractéristiques spectrales à long terme, sur des échantillons de plus de 10 s [BUNGE, 1977]. La classification est obtenue par une mesure de la distance de Mahalanobis et deux types de décision sont considérés : la classification en ensemble fermé et la discrimination. Les tests effectués sur un ensemble de 2500 énoncés provenant de 50 locuteurs ont

montré d'excellents résultats : un taux d'erreur de type I de 0,5% pour la classification et des taux d'erreur de type I et de type II de 1% pour la discrimination.

Malgré le développement de techniques de compensation des variations du canal de transmission téléphonique, le manque de robustesse de la méthode dans des conditions forensiques réelles a conduit à son abandon dans les années 1980 [BUNGE, 1979 ; BUNGE, 1991].

6.4.3. Computer Assisted Voice Identification System (CAVIS) - USA (1985 - 1989)

Le projet CAVIS a été développé par NAKASONE et MELVIN et cofinancé par le *Los Angeles County Sheriff's Department*, le *National Institute of Justice* et le *United States Secret Service*. Le but est de développer un système objectif, indépendant du texte et du canal de transmission. La base de données enregistrées pour l'expérimentation contient 10 échantillons de 30 s de parole spontanée provenant de 49 hommes blancs, enregistrés par téléphone, microphone et transmission radiophonique miniaturisée, lors de deux sessions distantes d'au moins deux mois. L'analyse de parole repose sur une numérisation du signal, échantillonné à 10,24 KHz et quantifié sur 12 bits, sur sa segmentation effectuée par un opérateur sur une base auditive et visuelle et sur une extraction des parties voisées du signal, de manière à constituer un échantillon d'au moins 10 s. Cinq paramètres sont extraits de ce signal dans le domaine temporel sur la base de transformées en ondelettes ; il s'agit de l'intensité totale des ondelettes, de la variation de cette intensité, de l'autocorrélation des ondelettes successives, de la fréquence fondamentale et de la distribution moyenne de l'énergie [NAKASONE ET MELVIN, 1989].

Dans le domaine spectral, le spectre entre 200 Hz et 2.45 KHz est divisé en neuf sous-bandes de même largeur spectrale et le spectre moyen à long terme normalisé par soustraction de la moyenne spectrale est extrait sous forme de neuf paramètres par transformée de Fourier rapide. Comme mesure d'intravariabilité, la stabilité des caractéristiques spectrales est calculée à l'intérieur d'une session et entre les deux sessions, ce qui permet de distinguer trois groupes de locuteurs ; ceux dont certaines caractéristiques restent stables durant les deux sessions, ceux pour lesquels cette stabilité est limitée à une seule session et ceux pour lesquels aucune stabilité ne peut être mise en évidence [NAKASONE ET MELVIN, 1989].

La procédure expérimentale a consisté à considérer les 245 (49 x 5) échantillons de parole enregistrés lors de la première session comme enregistrements inconnus et à les comparer aux 245 échantillons de la seconde session, considérés comme enregistrements de comparaison. Une analyse préliminaire consiste à calculer la stabilité des caractéristiques spectrales à l'intérieur de la seconde session et de rendre une « non-décision » lorsque celle-ci est faible ou inexistante. La métrique utilisée pour la comparaison combine la mesure d'une distance euclidienne pour les caractéristiques spectrales et une sommation des moindres carrés pour les caractéristiques temporelles. Le résultat est exploité soit dans une procédure de classification en ensemble fermé, soit dans une procédure de discrimination [NAKASONE ET MELVIN, 1989].

Les auteurs considèrent cependant que la distance ainsi mesurée doit plutôt être considérée comme un indice de proximité qu'une probabilité d'origine commune. Pour la procédure de classification, les résultats ne sont pas fournis uniquement lorsqu'à l'issue de la comparaison, l'enregistrement inconnu correspondant à l'enregistrement de comparaison est classé au premier rang, mais aussi lorsqu'il est classé au deuxième, au troisième, au septième et au quinzième rang. Les tests montrent que l'enregistrement inconnu est classé au premier rang dans 80% des cas, dans 85% des cas dans les deux premiers rangs, dans 91% des cas dans les trois premiers rangs, dans 95% des cas dans les sept premiers rangs et dans 99% dans les quinze premiers rangs [NAKASONE ET MELVIN, 1989].

Malgré des résultats intéressants obtenus à partir d'énoncés de parole enregistrés dans des conditions proches des conditions forensiques, le système CAVIS a été abandonné en 1992, sans avoir atteint le degré de fiabilité nécessaire à un usage dans des conditions réelles [KÜNZEL, 1994A ; NAKASONE, 1999].

6.4.4. Semi-AUtomatic Speaker Identification system (SAUSI) - USA (1976-1998)

Le système SAUSI a été développé par Harry HOLLIEN et ses collaborateurs entre 1976 et 1998, à l'*Institute for Advanced Study of the Communication Processes* de l'Université de Floride [DOHERTY, 1976 ; HOLLIEN ET JIANG, 1998]. Il a été utilisé pour la résolution de cas réels durant cette période, mais cette activité a cessé à la suite de la retraite de HOLLIEN.

Sur la base d'échantillons de parole d'une durée de 15 à 20 s, l'analyse du signal de parole repose sur quatre vecteurs de caractéristiques, mesurés sur des portions du signal de parole naturelle sélectionnées par un opérateur : (1) la fréquence fondamentale, mesurée par un système de bancs de filtres spécialement mis au point pour cette application, (2) le spectre de puissance moyen à long terme, (3) la mesure entre les trois premiers formants des voyelles et de la fréquence centrale de ces trois premiers formants, calculés par transformée de Fourier rapide et, finalement, (4) un vecteur regroupant des caractéristiques temporelles, comme le nombre de syllabes par unité de temps, les proportions entre temps de parole et temps de silence, entre temps d'élocution des consonnes et des voyelles [HOLLIEN, 1990].

La métrique utilisée pour la comparaison repose sur une mesure de la distance euclidienne pour chacun des quatre vecteurs de caractéristiques et sur une sommation de ces distances. L'évaluation des résultats est considérée dans une procédure de classification en ensemble ouvert : l'indice est comparé avec l'échantillon du locuteur suspect, mais aussi avec les échantillons d'un ensemble de six à dix autres locuteurs. La décision est laissée à l'appréciation de l'opérateur, qui fixe *a posteriori* et subjectivement selon l'ensemble des résultats à sa disposition un premier seuil, à partir duquel il prend une décision d'identification et un second seuil en deçà duquel il prend une décision d'exclusion ; aucune décision n'est rendue lorsque le résultat de la comparaison entre l'échantillon inconnu et l'échantillon du locuteur suspect se trouve entre ces deux seuils [HOLLIEN ET JIANG, 1998]. Malheureusement, aucune des nombreuses publications consultées concernant ce système ne propose une évaluation quantitative de ses performances.

6.4.5. IDentification Method (IDEM) – Italie (dès 1991)

Le système IDEM est développé par la *Fondazione Ugo Bordoni* (FUB) à Rome, depuis 1991. Ce logiciel semi-automatique se présente sous forme de modules indépendants, adaptés à la micro-informatique. La reconnaissance de locuteurs est basée sur la comparaison de la fréquence fondamentale et sur l'analyse des trois premiers formants des cinq voyelles /a/, /e/, /i/, /o/ et /u/. Le module d'acquisition du signal permet de numériser les signaux audio dans les formats informatiques habituels [FALCONE ET DE SARIO, 1994]. Ce système est utilisé en Italie pour la résolution de cas réels, seul ou conjointement avec d'autres méthodes propres à chaque expert.

Une écoute et un examen préliminaire, composés de la mesure du rapport signal sur bruit et du calcul du spectre de puissance à long terme, permet à l'opérateur d'évaluer la qualité du signal numérisé. Une durée minimale de 15 s de parole est nécessaire pour une expertise de reconnaissance de locuteurs. La segmentation est réalisée manuellement par l'opérateur à partir de la forme d'onde et d'une représentation spectrographique, afin de séparer les énoncés des interlocuteurs en cas de dialogue et d'extraire les voyelles et des portions stables du signal de parole. L'analyse repose sur une extraction manuelle de la fréquence fondamentale et des fréquences formantiques des voyelles, à partir d'une représentation graphique du spectre et du cepstre, calculés par transformée de Fourier rapide [FALCONE ET AL., 1995].

La comparaison et la décision d'identification reposent sur une mesure de la variabilité intralocuteur et interlocuteur des paramètres analysés, à l'aide de matrices de covariance et par l'application de tests statistiques. L'évaluation statistique repose sur la base de données téléphoniques *Speaker Identification and Verification Archives* (SIVA). Elle compte plus de 1000 locuteurs ayant appelé de toute l'Italie, pour la modélisation de la variabilité interlocuteur, et 40 locuteurs enregistrés à 20 reprises, pour la modélisation intralocuteur [FALCONE ET DE SARIO, 1994 ; PAOLONI, 1999].

6.4.6. REconnaissance Vocale Assistée par Ordinateur (REVAO) – France (1988 – 1993)

En 1989, le ministère français de l'Intérieur, par l'intermédiaire du Centre d'Études et de Recherche de la Police Nationale et de la Direction des Transmissions et de l'Informatique, a lancé un appel d'offres pour une « Étude, mise au point et présentation de moyens permettant une identification de locuteurs 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 » [BOË, 1998].

Devant les exigences extrêmes de ce cahier des charges aucun laboratoire en traitement de la parole, universitaire ou du Centre National de Recherche Scientifique (CNRS), n'y a répondu. En 1990, suite à cet appel d'offres, le Groupe Communication Parlée de la Société Française d'Acoustique (GFCP) a élaboré et adopté une motion précisant que l'identification de locuteurs est

un problème non résolu, que les méthodes utilisées ne sont pas fiables, et demandant que tout spécialiste se présentant comme expert en identification de locuteurs fasse la preuve de ses compétences avant de procéder à toute expertise ⁷⁹ [BOË, 1998].

Au moins un laboratoire a répondu à l'appel d'offres et a été retenu, la société Microsurfaces[®] sàrl de Besançon, exploitant des technologies développées à l'École Nationale Supérieure de Mécanique et de Microtechnique de Besançon (ENSMM) et collaborant avec l'Institut de Phonétique de Besançon. Dans leur projet de « recherche d'empreintes génétiques vocales », les auteurs proposent « de mettre en place les outils propres à un système permettant d'établir un 'portrait robot vocal' des individus, quels que soient leur langue, leur identité socioculturelle et leur état physiologique ».

En 1989, les premiers travaux de développement du logiciel de REcognition Vocale Assistée par Ordinateur (REVAO) entrepris par la société Microsurfaces[®] reposent sur l'application de la géométrie fractale de MANDELBROT [MANDELBROT, 1983]. Cette société avait déjà appliqué la théorie fractale dans le domaine forensique, en développant une méthode d'analyse de la dimension fractale de la surface des projectiles d'armes à feu en vue de l'identification de leur source. Aucun résultat probant n'a été obtenu par cette méthode, qui n'a fait l'objet d'aucune publication.

Le traitement du signal de parole est par essence différent de l'analyse de surfaces en trois dimensions, mais la société Microsurfaces[®] a considéré que l'approche fractale était tout de même appropriée. Leur méthode est basée sur l'hypothèse que les consonnes et les transitions entre phonèmes comportent des perturbations ayant une dimension fractale, comme l'explique un dossier publié dans la Revue de la Police Scientifique et Technique en 1991 : « Les consonnes possèdent un spectre renfermant des signaux caractéristiques du bruit du frottement de l'air contre les parois de la cavité buccale. Elles ont une dimension fractale susceptible d'être visualisée en fonction du temps et de la fréquence dans un espace tridimensionnel : le fractogramme » [ANONYME, 1991]. Cette représentation graphique ressemble au spectrogramme vocal, mais serait susceptible de révéler des caractéristiques spécifiques du locuteur permettant son identification par comparaison.

Malheureusement les résultats décevants obtenus par l'analyse fractale ou multifractale ont conduit les chercheurs à se tourner vers l'analyse du spectre à long terme, dont les limites concernant le pouvoir discriminatoire et la robustesse aux variations du canal de transmission étaient pourtant connues depuis longtemps au début des années 1990 ⁸⁰. Les travaux consécutifs à cette recherche n'ont fait l'objet d'aucune publication, le logiciel REVAO n'a pas été commercialisé. Les résultats obtenus ont été classifiés « confidentiel – défense » par le ministère français de l'Intérieur ; suite à une liquidation judiciaire, la société Microsurfaces[®] a disparu en 1993 après que son matériel informatique eut été mis sous séquestre [BOË, 1998].

⁷⁹ *supra* : 3.2.1. Le refus de témoigner

⁸⁰ *supra* : 6.3.1.1.2. Application

Les raisons de cette classification sont obscures, mais les informations et les événements, rendus publics alors, laissent à penser qu'elle a servi à éviter aux autorités mandantes un discrédit devant l'ampleur d'un échec. Cette décision de classification est d'autant plus regrettable qu'elle sert à justifier l'efficacité de cette méthode et le bien-fondé de son utilisation par la Police Nationale française encore actuellement, sans aucune démonstration scientifique évidemment, puisque la méthode est secrète. C'est en tout cas ce que sous-entend une circulaire de présentation du laboratoire d'analyse et de traitement de signal à destination des autorités judiciaires et policières françaises et suisses romandes : « Les outils utilisés sont des outils mathématiques, l'informatique permettant une étude statistique des constantes du signal vocal étudié. Cette méthode a fait l'objet d'une classification en 'CONFIDENTIEL DEFENSE' et fait suite à une étude décidée par le ministère de l'Intérieur en 1989. Il convient de préciser que cette approche suscite le plus grand intérêt chez les industriels ».

A la fin de l'année 1997, le GFCP a réitéré sa motion de 1990, suite à une expertise controversée d'identification de locuteurs, effectuée notamment par le laboratoire d'analyse et de traitement de signal de la Police Nationale française [BOË, 1998]. La seule manière de rendre l'utilisation du logiciel REVAO acceptable, d'un point de vue scientifique, est de proposer un protocole d'évaluation de cette méthode admis par tous les partenaires.

6.4.7. Approches récentes

Dans son rapport « *Voice Analysis* » présenté au congrès de l'Interpol en 1998, BRAUN mentionne la tendance actuelle à concentrer les efforts sur des procédures plus objectives et moins gourmandes en temps de travail [BRAUN, 1998]. Cette tendance s'observe par une activité de publication dans le domaine de la reconnaissance automatique de locuteurs en sciences forensiques, en forte progression depuis 1997.

D'une part, plusieurs systèmes de reconnaissance automatique ou semi-automatique en activité, provenant surtout d'Europe de l'Est, ont été décrits pour la première fois dans la littérature internationale et d'autre part plusieurs laboratoires académiques ou de police décrivent l'avancement de leurs recherches dans la mise au point de systèmes automatiques ou semi-automatiques en vue d'une application forensique.

Le système semi-automatique « DIALECT », mis au point par le laboratoire forensique du service de sécurité fédéral d'ex-Union Soviétique, est actuellement utilisé par les ministères de l'Intérieur de Russie et d'Ukraine. Il est composé d'un module d'analyse indépendant du texte basé sur un vecteur de 378 paramètres, d'un module d'analyse des caractéristiques de la fréquence fondamentale basé sur un vecteur de 123 paramètres et d'un module d'analyse des quatre voyelles russes les plus fréquentes, analysées sur la base d'un vecteur de 144 paramètres. Le module de décision statistique s'appuie sur la détermination de seuils par l'estimation expérimentale des fonctions de distribution des paramètres dans la variabilité intralocuteur et interlocuteur [TIMOFEEV ET SIMAKOV, 1998 ; BRAUN, 1998].

Le département d'examen « phonoscopique » de l'Institut d'expertises forensiques de Lituanie décrit le système *Speaker Identification by the Voice* (SIVE), basé sur des coefficients de prédiction linéaire et des coefficients cepstraux extraits des parties pseudo-stationnaires du signal de parole. La classification est effectuée soit par un classificateur gaussien, soit par quantification vectorielle. Ce système est utilisé depuis 1991 et est aussi en fonction au ministère de l'Intérieur de Pologne [LIPEIKA ET LIPEIKIENE, 1996 ; BRAUN, 1998].

L'étude de la variabilité intralocuteur est au centre des recherches menées par le département d'informatique de l'Université d'Etat de Caroline du Nord. Les buts de cette recherche de grande envergure sont centrés autour de la voix déguisée : la détection automatique du déguisement et l'évaluation des performances de plusieurs types de classificateurs comme les modèles de Markov cachés, la quantification vectorielle ou les réseaux neuromimétiques, en fonction du type de déguisement [RODMAN, 1998]. Le département « acoustique et signal » de l'Institut de Recherche Criminelle de la Gendarmerie Nationale française (IRCGN) a conduit une étude visant à quantifier l'influence des émotions dans la variabilité intralocuteur et développé une méthode particulière d'extraction robuste du contour de F_0 [MARESCAL, 1999].

Pour évaluer la variabilité intralocuteur en espagnol castillan, la police judiciaire espagnole a enregistré la base de données « AHUMADA », en collaboration avec l'Université Polytechnique de Madrid. Ce corpus est testé à l'aide d'un système de vérification, en vue d'une application forensique. Les paramètres analysés sont des coefficients cepstraux en échelle Mel (MFCC) et la mesure de similarité est assurée par un classificateur par mélanges de fonctions de densité gaussiennes (GMM). Les auteurs reportent des taux d'erreur de 8,5% lorsque les modèles et les tests proviennent de sessions différentes et sont constitués de parole spontanée [ORTEGA-GARCIA ET AL., 1998].

Le choix d'une méthode entièrement automatique est aussi opéré par d'autres laboratoires de police. Le laboratoire de police scientifique de Tokyo a développé un système d'identification en ensemble ouvert à l'aide d'un réseau neuromimétique exploitant des fréquences formantiques extraites de la parole continue. Il mentionne un taux d'identification correcte sur une base de données de 50 locuteurs [BRAUN, 1998].

En collaboration avec le ministère des affaires intérieures d'Ukraine, l'Académie des Sciences d'Ukraine développe le système *Crime-detection Automatic Speaker Verification and Identification* (CASVI). Les auteurs ne détaillent pas la méthode, mais reportent des taux d'identification de 90% sur la base de signaux dont le rapport signal sur bruit est de 12 dB [GORBAN ET AL. ; 1999].

Le laboratoire national des forces de gendarmerie turques développe depuis 1994 un système d'identification de locuteurs semi-automatique, nommé KASIS. Il permet l'extraction de paramètres tels que F_0 , les fréquences formantiques, des coefficients de prédiction linéaire, des coefficients cepstraux en échelle Mel, ainsi qu'une représentation spectrographique. Le module de décision est entièrement subjectif ; l'examineur se forge une opinion sur la base des ressemblances et des différences qu'il observe entre les différents paramètres analysés et fournit

une conclusion sur le modèle de l'échelle de conclusions préconisée par le *Voice Identification and Acoustic Analysis Subcommittee* (VIAAS) de l'*International Association for Identification* (IAI)⁸¹.

6.5. Conclusion

La reconnaissance automatique de locuteurs suscite l'intérêt des chercheurs, des industriels et des criminalistes depuis presque quarante ans. Un certain consensus existe autour des caractéristiques dépendantes du locuteur et des classificateurs les plus efficaces, notamment grâce au test d'évaluation mis au point annuellement depuis 1995 par le *National Institute of Standards* nord-américain (NIST) [FURUI, 1997 ; PRZYBOCKI ET MARTIN, 1998].

Les résultats de ce test indiquent de toute évidence que la technologie n'a pas encore atteint un niveau de maturité permettant son utilisation à large échelle ni dans le domaine commercial ni dans le domaine forensique, malgré des progrès constants et significatifs [PRZYBOCKI ET MARTIN, 1998 ; BOVES, 1998]. De plus, la recherche fondamentale semble un peu délaissée dans le contexte économique actuel, qui privilégie les synergies avec le monde industriel, alors que seule une approche de ce problème avec un regard nouveau permettrait un saut qualitatif significatif des performances. NOLAN mentionne notamment qu'une analyse sous l'angle phonologique est ostensiblement absente du domaine de la reconnaissance de locuteurs [NOLAN, 1995].

Finalement, les applications commerciales d'identification de personnes basées sur l'analyse du signal de parole sont aujourd'hui concurrencées par des systèmes exploitant d'autres mesures biométriques, comme l'empreinte digitale ou le réseau vasculaire rétinien, dont l'intravariabilité est faible ou nulle. Les applications pour lesquelles il n'existe aucune alternative à la reconnaissance de locuteurs sont rares ; les applications téléphoniques en font partie [BOVES, 1998].

Dans le domaine forensique, l'expertise en reconnaissance de locuteurs est souvent considérée comme une « *ultima ratio* », lorsque toutes les autres voies d'investigation ont été épuisées ou lorsque la voix enregistrée représente le seul lien entre l'auteur et l'infraction. Dans les années 1970, un effort tout particulier a été consenti au niveau de la recherche fondamentale, surtout par les administrations nord-américaine et allemande, avec pour résultats le développement de systèmes à la pointe du progrès, grâce à la collaboration avec des partenaires industriels.

La difficulté de la tâche et les échecs successifs de ces systèmes, lors de leur application dans des conditions forensiques réelles, a conduit les industriels à se tourner vers des applications moins complexes, comme le contrôle d'accès, et les administrations à s'engager dans des domaines de recherche plus gratifiants. Les échecs des années 1970 peuvent être partiellement expliqués par le manque de robustesse des méthodes utilisées, par la qualité technique catastrophique des enregistrements soumis pour analyse et par le développement de modules de décision focalisés sur des décisions binaires d'acceptation ou de rejet.

⁸¹ *supra* : 5.4.3. Les standards de l'IAI

La place toujours plus importante de la téléphonie mobile dans les activités humaines, licites ou illicites, est certainement une des causes du regain d'intérêt pour la reconnaissance de locuteurs de la part des acteurs du monde judiciaire. Les recherches récentes montrent que les laboratoires utilisent maintenant généralement des méthodes d'analyse et de classification qui représentent l'état de l'art dans le domaine de la reconnaissance de locuteurs.

Par contre, le processus d'inférence de l'identité du locuteur est toujours perçu en terme de discrimination ou de classification. Cette constatation montre encore une fois que l'effort principal est porté sur l'outil d'analyse et sa maîtrise, alors que peu de réflexion est accordée à l'interprétation de l'information fournie par cet outil dans un cadre forensique.

Ce point de vue conduit à la mise en place de processus d'évaluation et de validation focalisés sur des décisions binaires qui ne rendent pas compte de manière satisfaisante des performances des systèmes testés. Par exemple, si un système de classification classe toujours la vraie source parmi les cinq meilleurs candidats sur mille candidats possibles, mais seulement dans 20% des cas au 1^{er} rang, ce système ne sera crédité que d'un taux de classification correcte de 20%, alors que l'information délivrée par le système est d'excellente qualité.

Cet exemple montre que le choix d'un type d'inférence de l'identité inadapté à la problématique forensique a pour grave conséquence de discréditer de manière abrupte une approche automatique de la reconnaissance de locuteurs en rendant compte de manière très imparfaite de ses performances.

PARTIE 3

RECHERCHE EXPERIMENTALE

VII. DEVELOPPEMENT D'UN SYSTEME AUTOMATIQUE DE RECONNAISSANCE DE LOCUTEURS

7.1. Introduction

L'objectif de cette partie expérimentale est double. Le premier but consiste à développer un système de reconnaissance de locuteurs basé sur des méthodes d'analyse et de classification représentant l'état de l'art dans le domaine de la reconnaissance automatique de locuteurs. Le second, développé dans le chapitre VIII, consiste à évaluer l'outil réalisé dans un cadre bayésien, à l'aide de bases de données d'énoncés de parole, dont la qualité n'est pas supérieure à celle qui peut être atteinte lors de l'enregistrement d'un message anonyme ou d'une écoute téléphonique.

7.2. Le système de reconnaissance de locuteurs

7.2.1. Définition générale du système

7.2.1.1. Choix de la méthode d'analyse du signal de parole

Les méthodes d'analyse actuelles sont basées sur la variabilité implicite du signal de parole en fonction du locuteur. Les caractéristiques spectrales du signal dépendantes du locuteur sont extraites soit par la méthode de prédiction linéaire soit par l'analyse homomorphique. En plus de leur efficacité, ces méthodes ont l'avantage de réduire le temps et la complexité du travail de l'opérateur puisque toute segmentation manuelle est inutile. Cette approche limite par là même l'influence de la subjectivité humaine dans le processus analytique.

Le choix s'est porté sur la prédiction linéaire perceptuelle (PLP)⁸², qui est d'une part l'une des méthodes les plus couramment utilisées et d'autre part une méthode peu affectée par des différences de niveau sonore des signaux analysés et dont la robustesse aux différentes dégradations est plus universelle que d'autres méthodes d'extraction des coefficients de prédiction linéaire [YEGNANARAYANA *ET AL.*, 1992 ; OPENSHAW *ET AL.*, 1993 ; RAMACHANDRAN *ET AL.*, 1995]. Par contre, aucune méthode de compensation de l'effet du canal de transmission n'a été utilisée, pour définir de manière claire les limites du système et pour permettre de quantifier un apport ultérieur de techniques de compensation, par exemple basées sur la théorie des paramètres manquants [EL MALIKI ET DRYGAJLO, 1998].

7.2.1.2. Choix du classificateur

Pour une application forensique, le choix du classificateur dépend principalement de sa capacité à fonctionner en mode indépendant du texte et à fournir une mesure de similarité sous la

⁸² *supra* : 6.2.3.3.2. Paramètres dérivés de la prédiction linéaire

forme d'un nombre réel, faisant partie d'un ensemble de données continues. Plusieurs méthodes remplissent cette condition, mais la modélisation par mélange de fonctions de densité gaussiennes (GMM) représente l'état de l'art pour la reconnaissance de locuteurs en mode indépendant du texte, lorsque la quantité de données nécessaire à la constitution du modèle est suffisante.

Cette dernière condition est remplie dans le domaine forensique, puisque le modèle statistique de la voix des locuteurs est réalisé soit à partir des sessions d'enregistrement d'une base de données, pour la modélisation de la variabilité interlocuteur, soit à partir des enregistrements de comparaison, pour la modélisation de la variabilité intralocuteur des personnes mises en cause. L'indice matériel est toujours considéré comme enregistrement de test, notamment par le fait que la maîtrise de ses caractéristiques n'est que très partielle lors de la collecte de l'élément de preuve.

La mesure de similarité, calculée par la méthode GMM, est la probabilité conditionnelle des caractéristiques de l'enregistrement de test, sachant celles du modèle [REYNOLDS ET ROSE, 1995]. Ce résultat est un nombre réel et représente l'élément de preuve E utilisé pour l'évaluation du rapport de vraisemblance.

7.2.2. Architecture du système

L'architecture du système repose sur quatre modules logiciels, « SILREM », « PLP », « GMM » et « GMM-evalue », implémentés en langage C par Monsieur Mounir El MALIKI, doctorant au laboratoire de traitement des signaux (LTS) de l'École Polytechnique Fédérale de Lausanne (EPFL). Le système ainsi constitué remplit deux fonctions distinctes : durant la phase d'entraînement, il permet de constituer des modèles statistiques de la voix des locuteurs avec le module « GMM », à partir de données d'entraînement (Figure VII.1.). Durant la phase de test, il permet de comparer les modèles réalisés à des enregistrements de test, avec le module « GMM-evalue ».

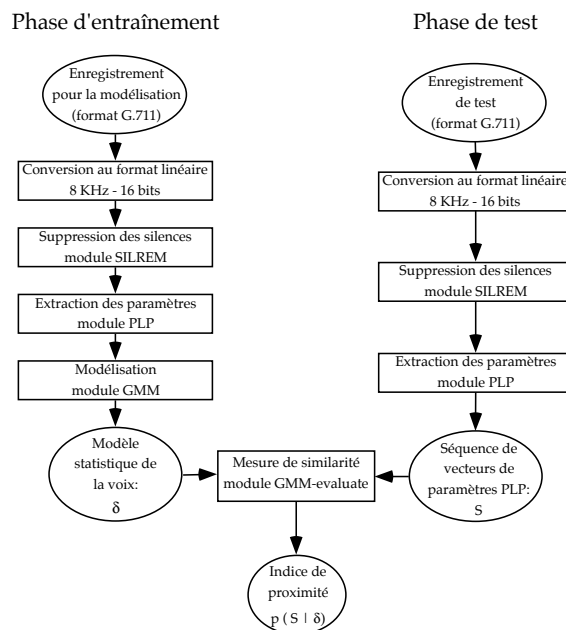


Figure VII.1. Architecture du système de reconnaissance de locuteurs

7.2.3. Prétraitement du signal

7.2.3.1. Suppression des silences

Le module de suppression des silences « SILREM » est basé sur l'algorithme de Murphy décrit par REYNOLDS : Un seuil adaptatif permet de séparer les endroits de forte énergie, considérés comme de la parole, des endroits de faible énergie, considérés comme des silences [REYNOLDS, 1992].

7.2.3.2. Paramétrisation

Le module de paramétrisation « PLP » permet d'extraire les coefficients de prédiction linéaire perceptuels, selon la méthode de HERMANSKY [HERMANSKY, 1990]. Un vecteur de douze paramètres PLP est extrait de chaque fenêtre de 20 ms. Le fenêtrage du signal est de type « Hamming »⁸³ et le recouvrement des fenêtres est de 10 ms (Figure VII.2.).

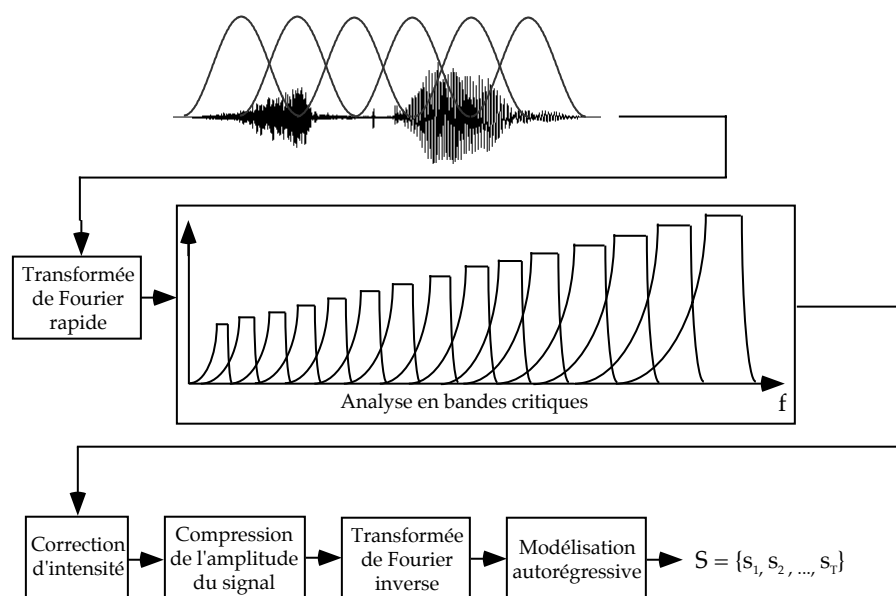


Figure VII.2. Méthode d'extraction des paramètres de prédiction linéaire perceptuelle PLP [HERMANSKY, 1990]

7.2.3.3. Modélisation

Cette étape consiste à modéliser la distribution des douze paramètres PLP (Figures VII.3. et VII.4.). Chaque paramètre est modélisé par un mélange de M fonctions de densité gaussiennes, définies par leur moyenne μ et leur variance σ^2 . Chaque locuteur est représenté par un modèle GMM δ , issu de cette modélisation. Le modèle δ est constitué des vecteurs de moyennes $\bar{\mu}_i$, des vecteurs de variance extraits de la diagonale de la matrice de covariance Σ_i et des facteurs r_i , qui pondèrent l'importance de chaque fonction de densité gaussienne dans le modèle δ .

$$\delta = \{r_i, \bar{\mu}_i, \Sigma_i\} \quad i = 1, \dots, M \quad (7.1)$$

⁸³ *supra*: 6.2.2.2.2. Analyse du spectre à court terme par transformée de Fourier

Dans le cadre de cette recherche, les enregistrements de parole utilisés pour la modélisation durent de 80 s à 140 s. Pour ce type de durée, les essais réalisés par le concepteur de l'algorithme montrent que la modélisation des paramètres obtenue avec un mélange de 64 fonctions de densité gaussiennes est la plus efficace ; ce nombre de gaussiennes a donc été retenu.

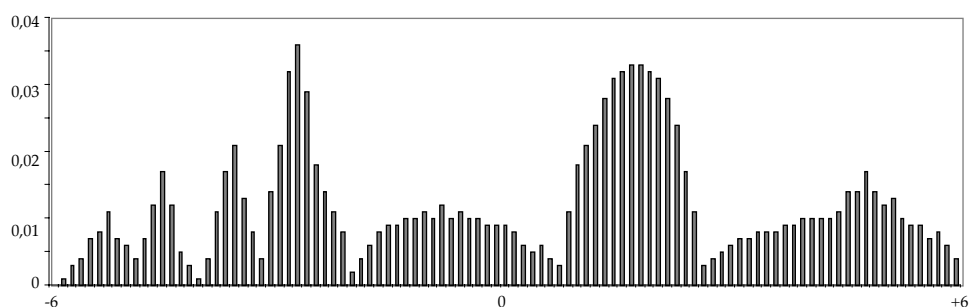


Figure VII.3. Histogramme simulé de la distribution d'un seul paramètre PLP

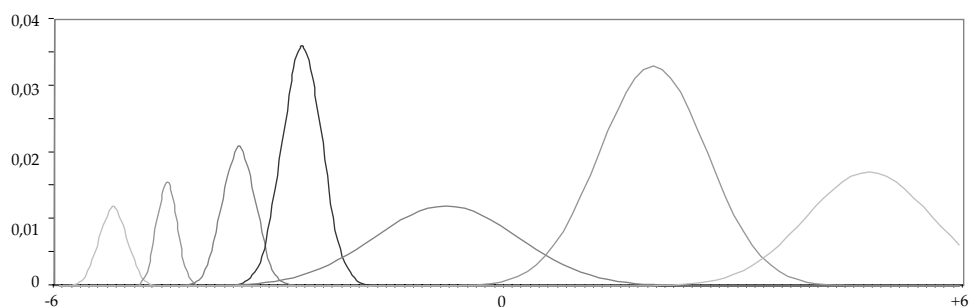


Figure VII.4. Exemple de modélisation de la distribution par un mélange de sept fonctions de densité gaussiennes

Le module logiciel « GMM » permet d'estimer les paramètres du modèle δ , de manière à ce qu'il corresponde le mieux possible aux données d'entraînement. Plusieurs techniques ont été développées pour l'entraînement du modèle GMM, mais la plus utilisée est l'estimation de la vraisemblance maximale, *maximum likelihood* (ML) [REYNOLDS ET ROSE, 1995]. Son but est de déterminer les paramètres qui maximisent la vraisemblance du modèle, sur la base d'une séquence de n vecteurs d'entraînement $Z = \{\bar{z}_1, \dots, \bar{z}_n\}$.

Comme cette maximisation ne peut pas être calculée directement, la modélisation est réalisée de manière itérative par l'algorithme *Expectation Maximisation*⁸⁴. Elle débute avec un modèle initial (δ), qui est utilisé pour estimer un nouveau modèle $\bar{\delta}$, de manière à ce que $p(\bar{z}_i | \bar{\delta}) \geq p(\bar{z}_i | \delta)$. Le nouveau modèle devient le modèle initial pour l'itération suivante et la procédure est répétée jusqu'à ce que le seuil de convergence désiré soit atteint. Cette procédure est similaire à la technique utilisée pour entraîner les modèles de Markov cachés⁸⁵, avec l'algorithme de réestimation de Baum-Welch [BAUM ET AL., 1970 IN : REYNOLDS ET ROSE, 1995]. Dans le cadre de

⁸⁴ *supra* : 6.3.2.3. Modélisation par mélange de fonctions de densité gaussiennes

⁸⁵ *supra* : 6.3.2.4. Modélisation par modèles de Markov cachés

cette recherche, chaque modèle a été soumis à une procédure de 30 itérations, pour assurer sa convergence sur les données.

7.2.3.4. Comparaison

Le module logiciel « GMM-evaluate » permet de comparer un modèle GMM (δ) à une séquence de vecteurs de paramètres $S = \{\bar{s}_1, \bar{s}_2, \dots, \bar{s}_T\}$, calculée avec le module « PLP », à partir d'un enregistrement de test. La comparaison consiste à calculer la probabilité conditionnelle des vecteurs, sachant le modèle, $p(S | \delta)$, dans l'hypothèse où les vecteurs (\bar{s}_t) sont indépendants.

$$p(S | \delta) = \prod_{t=1}^T p(\bar{s}_t | \delta) \quad (7.2)$$

avec

$$p(\bar{s}_t | \delta) = \sum_{i=1}^M r_i b_i(\bar{s}_t) \quad (7.3)$$

Chaque fonction de densité qui compose le modèle δ est une fonction gaussienne b_i , exprimée en fonction du vecteur \bar{s}_t de dimension D , du vecteur de moyennes $\bar{\mu}_i$, de la matrice de covariance diagonale Σ_i et du facteur de pondération r_i :

$$b_i(\bar{s}_t) = \left(\frac{1}{\sqrt{\Sigma_i}^D \sqrt{(2\pi)^D}} \right) \exp \left\{ -\frac{(\bar{s}_t - \bar{\mu}_i)^2}{2 \Sigma_i} \right\} \quad \text{avec} \quad \sum_{i=1}^M r_i = 1 \quad (7.4)$$

La comparaison consiste à mesurer le vecteur \bar{s}_t dans chaque fonction de densité de probabilité b_i du modèle δ (Figure VII.5.). Ces probabilités, considérées comme indépendantes, sont multipliées entre elles pour calculer la vraisemblance de l'enregistrement de test par rapport au modèle, $p(S | \delta)$.

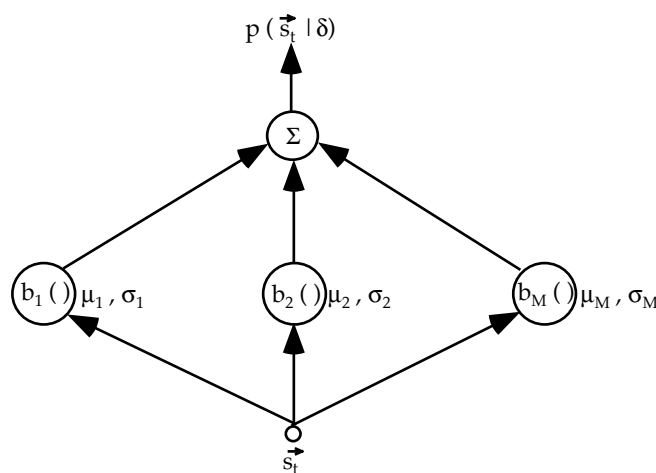


Figure VII.5. Calcul de la vraisemblance du vecteur de paramètres \bar{s} par rapport au modèle GMM δ composé de M fonctions de densité de probabilité gaussiennes

7.3. Méthode de calcul du rapport de vraisemblance

7.3.1. Production des données

7.3.1.1. Estimation de la distribution de la variabilité intralocuteur

Pour un locuteur Y , l'estimation de la variabilité intralocuteur est obtenue par la comparaison de l'ensemble $M = \{\delta_{Y\omega} \dots, \delta_{Yi}\}$ des modèles de sa propre voix, avec l'ensemble $C = \{S_{Y\omega} \dots, S_{Y\phi}\}$ des enregistrements de comparaison de la voix de ce même locuteur. Les modèles de l'ensemble M sont calculés avec le module logiciel GMM, les séquences de vecteurs de paramètres sont extraits des enregistrements de comparaison par le module logiciel « PLP » et la comparaison est réalisée avec le module logiciel « GMM-evaluate ». La comparaison de tous les éléments de l'ensemble M avec tous les éléments de l'ensemble C , de $p(S_{Y\alpha} | \delta_{Y\alpha})$ à $p(S_{Yi} | \delta_{Y\phi})$ permet d'obtenir des scores, sous forme d'un ensemble de nombres réels. Cet ensemble de données $A = \{a_1 ; \dots ; a_m\}$, décrit la variabilité intralocuteur du locuteur Y .

7.3.1.2. Estimation de la distribution de la variabilité interlocuteur

L'estimation de la variabilité interlocuteur de l'indice matériel X est obtenue par la comparaison des vecteurs de paramètres de cet échantillon S_x , avec les modèles des voix de l'ensemble $P = \{\delta_{P\omega} \dots, \delta_{Pk}\}$ des personnes qui modélisent la population potentielle des auteurs de l'indice matériel X . Les modèles de l'ensemble P sont calculés avec le module logiciel GMM, la séquence de vecteurs de paramètres est extraite de l'indice matériel X par le module logiciel « PLP » et la comparaison est réalisée avec le module logiciel « GMM-evaluate ». La comparaison de l'indice matériel X avec tous les éléments de l'ensemble P , de $p(S_x | \delta_{P\omega})$ à $p(S_x | \delta_{Pk})$ permet d'obtenir des scores, sous forme d'un ensemble de nombres réels. Cet ensemble de données $B = \{b_1 ; \dots ; b_n\}$, décrit la variabilité interlocuteur de l'indice matériel X .

7.3.2. Distribution des données

Dans un premier temps la distribution des données A et B , issues de l'estimation des variabilités intralocuteur et interlocuteur, a été approchée par une simple fonction gaussienne [MEUWLY ET AL., 1998]. Cependant, s'il existe des cas où la distribution des données A et B suit une fonction de densité de probabilité gaussienne, leur distribution est la plupart du temps multimodale ou asymétrique et ne peut être estimée par aucune loi de distribution connue, comme le montrent quelques exemples provenant de locuteurs de la base de données « Polyphone IPSC »⁸⁶ (Figure VII.6). Dans un deuxième temps, l'estimation a été réalisée de manière plus précise, par *kernel density estimation*.

⁸⁶ *infra* : 8.2.2.2. Composition de la base de données « Polyphone ISPC »

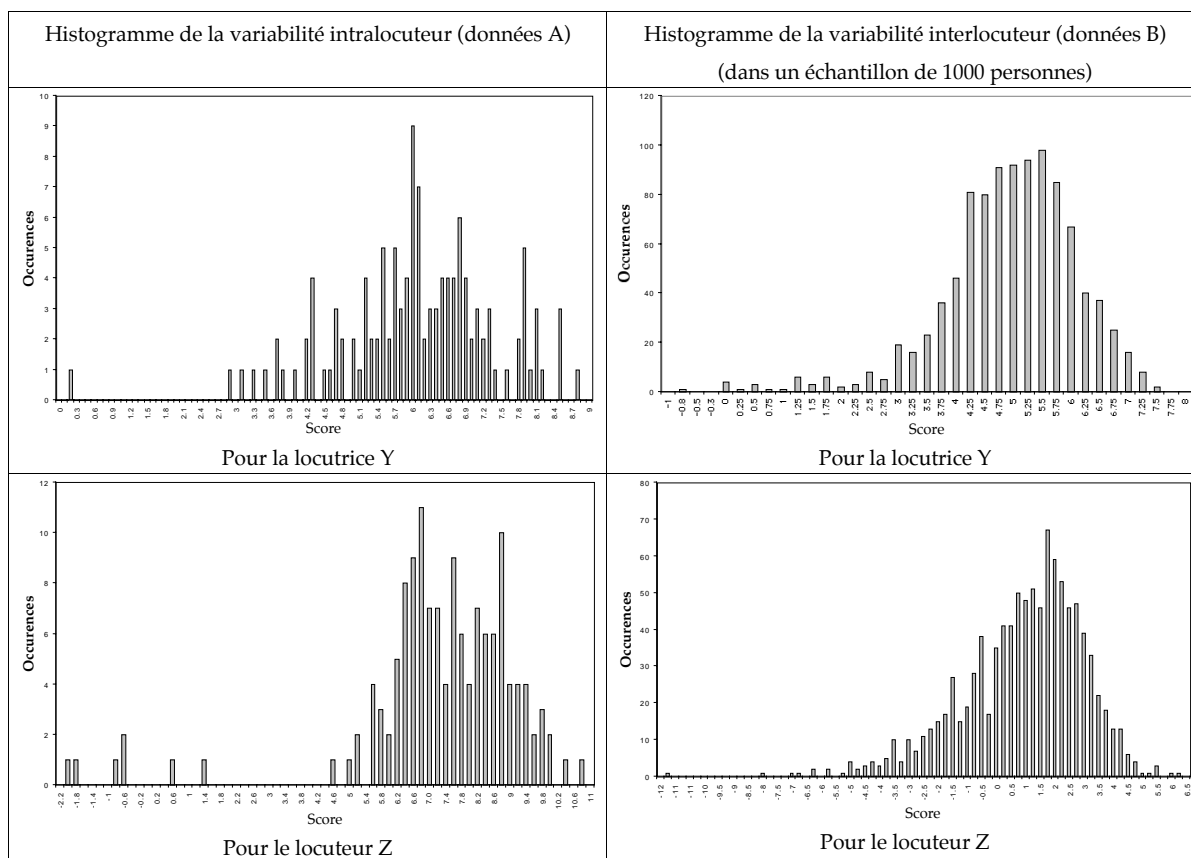


Figure VII.6. Distribution des indices de proximité calculés par le classificateur GMM

7.3.3. Estimation de la distribution par *kernel density estimation*

Une idée peu répandue consiste à considérer les données elles-mêmes comme source de la fonction de densité de probabilité [AITKEN, 1995]. L'estimation de cette densité de probabilité n'est pas trop difficile si la distribution des données est suffisamment lisse. En sciences forensiques, AITKEN a proposé l'application de l'estimation de la densité par noyau, *kernel density estimation*, décrite par SILVERMAN [SILVERMAN, 1986 ; AITKEN, 1995].

Cette méthode peut être considérée comme un développement de l'histogramme. Dans l'estimation de la densité par noyau, les blocs rectangulaires correspondant à une observation dans l'histogramme sont remplacés par une fonction noyau, *kernel function*, en général une courbe de densité de probabilité gaussienne, centrée sur l'observation qu'elle décrit. L'estimation de la courbe de densité de probabilité est ensuite obtenue en additionnant l'ensemble des courbes qui décrivent les observations et en divisant cette somme par le nombre d'observations. Comme chaque composante de la somme est une fonction de densité de probabilité, chacune a une aire égale à 1. La division par le nombre d'observations permet d'obtenir une aire de 1, sous la courbe d'estimation de la distribution des données, et d'en faire ainsi une fonction de densité de probabilité $f(\theta)$ [AITKEN, 1995].

Dans l'histogramme, la précision de la description de la distribution est conditionnée par la largeur des intervalles. Dans l'estimation de la densité par noyau, elle est conditionnée par la variance des courbes de densité de probabilité gaussiennes, qui peut être calculée à partir de l'ensemble des données F , \bar{z} représentant la moyenne des mesures :

$$\sigma^2 = \sum_{i=1}^k \frac{(z_i - \bar{z})^2}{k-1} \quad F = \{z_1, \dots, z_k\} \quad (7.5)$$

L'écart type σ est ensuite multiplié par un paramètre de lissage λ , *smoothing parameter*, qui détermine le lissage de la courbe ; la valeur du paramètre a été fixée à la valeur moyenne de 0.5, valeur recommandée par le logiciel de traitement statistique S plus®. Pour chaque noyau, la fonction de densité de probabilité $K(\theta | z_i, \lambda)$ suit une loi de distribution gaussienne de moyenne z_i et de variance $\lambda^2 \sigma^2$:

$$K(\theta | z_i, \lambda) = \left(\frac{1}{\lambda \sigma \sqrt{2\pi}} \right) \exp \left\{ -\frac{(\theta - z_i)^2}{2\lambda^2 \sigma^2} \right\} \quad (7.6)$$

L'estimation de la fonction de densité des données F , $\hat{f}(\theta | D, \lambda)$, est calculée de la manière suivante :

$$\hat{f}(\theta | F, \lambda) = \frac{1}{k} \sum_{i=1}^k K(\theta | z_i, \lambda) \quad (7.7)$$

7.3.4. Estimation des fonctions de densité de probabilité

Pour modéliser le cas où l'hypothèse H_1 ⁸⁷ est vérifiée, la fonction de densité de probabilité est estimée par *kernel density estimation*, à partir des données qui décrivent la variabilité intralocuteur (A). Elle est calculée de la manière suivante (Figure VII.7.):

$$\hat{f}(\theta|A, \lambda) = \frac{1}{m} \sum_{i=1}^m K(\theta|a_i, \lambda) \quad A = \{a_1, \dots, a_m\} \quad (7.8)$$

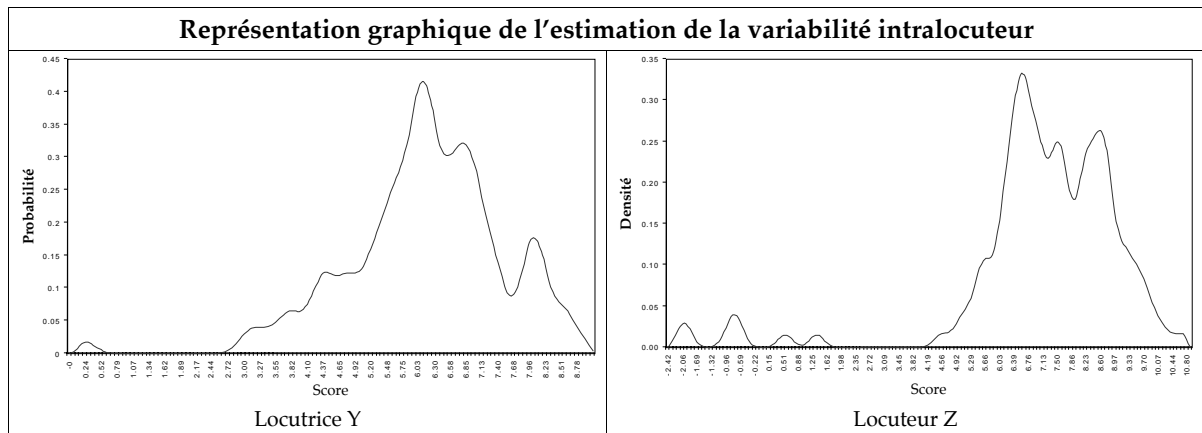


Figure VII.7. Estimation de la variabilité interlocuteur par *kernel density estimation*

Pour modéliser le cas où l'hypothèse H_2 ⁸⁸ est vérifiée, la fonction de densité de probabilité est estimée par *kernel density estimation*, à partir des données qui décrivent la variabilité interlocuteur (B). Elle est calculée de la manière suivante (Figure VII.8) :

$$\hat{f}(\theta|B, \lambda) = \frac{1}{n} \sum_{i=1}^n K(\theta|b_i, \lambda) \quad B = \{b_1, \dots, b_n\} \quad (7.9)$$

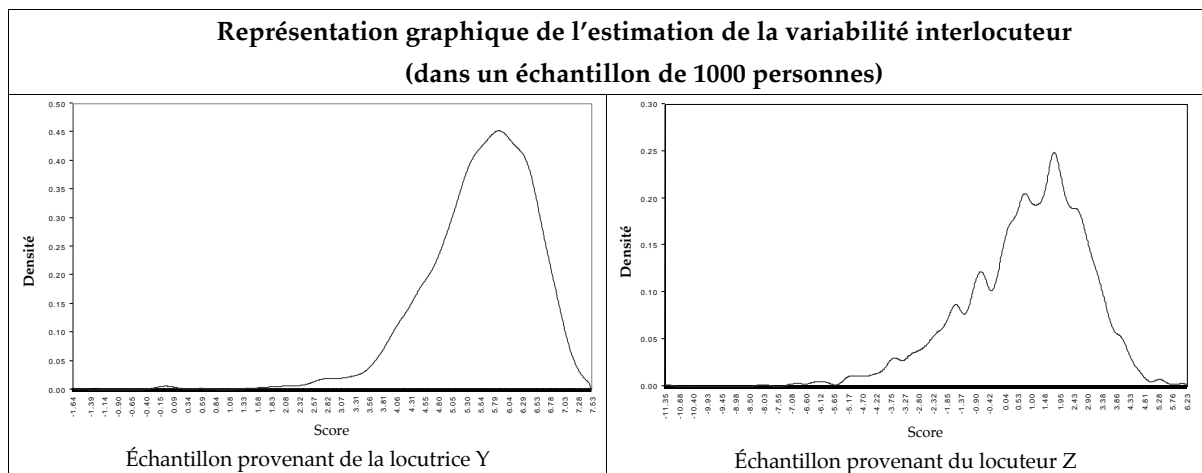


Figure VII.8. Estimation de la variabilité interlocuteur par *kernel density estimation*

⁸⁷ *supra*: 3.5.5.1.5. Formalisation

⁸⁸ *supra*: 3.5.5.1.5. Formalisation

Lorsque l'estimation de la densité au point E est égale à zéro, elle est remplacée par une densité estimée de $1 * 10^{-89}$. Cette attitude certainement conservatrice a été choisie en fonction de la taille de la base de données utilisée pour la modélisation de la population potentielle ; en effet dans ce cas, la valeur E représente la seule occurrence parmi les 10^3 personnes que contient cette base de données.

7.3.5. Calcul du rapport de vraisemblance de l'élément de preuve E

7.3.5.1. Vraisemblance de l'élément de preuve E lorsque H_1 est vraie

Dans le cas où l'hypothèse H_1 est vérifiée, la vraisemblance de l'élément de preuve E, $p(E | H_1)$ ⁹⁰, est déterminée dans la fonction de densité de probabilité de la variabilité intralocuteur, à la valeur E. Elle est calculée de la manière suivante :

$$\hat{f}(E | A, \lambda) = \frac{1}{m} \sum_{i=1}^m K(E | a_i, \lambda) \quad (7.10)$$

Par exemple, dans l'hypothèse où l'indice X provient de la locutrice Y, la vraisemblance d'un élément de preuve valant 6 est estimée à 0,40 (Figure VII.9.). Dans l'hypothèse où l'indice X provient du locuteur Z, la vraisemblance d'un élément de preuve valant 6 est estimée à 0,15 (Figure VII.9.).

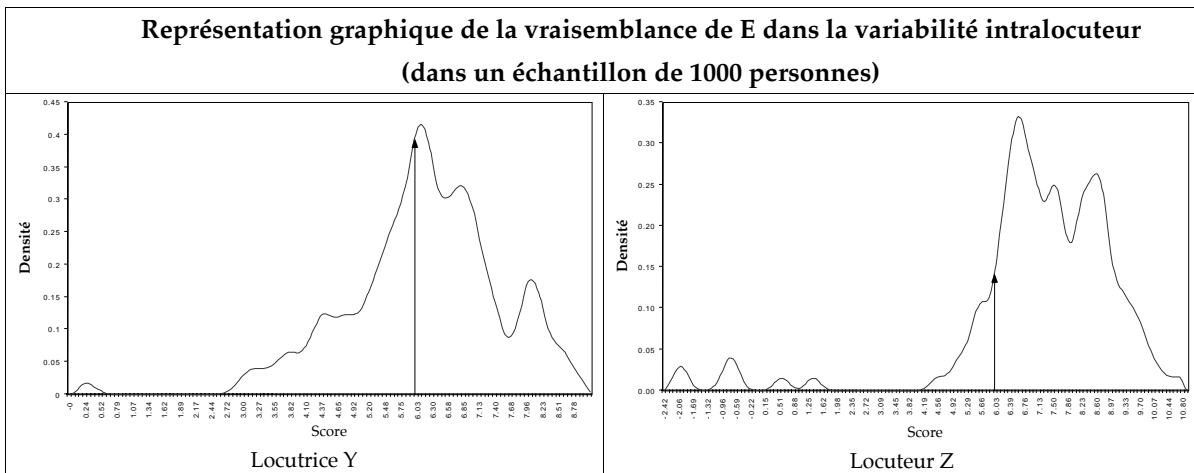


Figure VII.9. Calcul de la vraisemblance de E valant 6, dans le cas où l'hypothèse H_1 est vérifiée

⁸⁹ Cette approche néglige la contribution de ces valeurs plancher à l'intégrale de la densité.

⁹⁰ *supra*: 3.5.5.1.5. Formalisation

7.3.5.2. Vraisemblance de l'élément de preuve E lorsque H_2 est vraie

Dans le cas où l'hypothèse H_2 est vérifiée, la vraisemblance de l'élément de preuve E, $p(E | H_2)$ ⁹¹, est déterminée dans la fonction de densité de probabilité de l'hypothèse H_2 , à la valeur E. Elle est calculée de la manière suivante :

$$\hat{f}(E | B, \lambda) = \frac{1}{n} \sum_{i=1}^n K(E | b_i, \lambda) \quad (7.11)$$

Par exemple, dans l'hypothèse où l'indice X ne provient pas de la locutrice Y, la vraisemblance d'un élément de preuve valant 6 est estimée à 0,22 (Figure VII.10.). Dans l'hypothèse où l'indice X ne provient pas du locuteur Z, la vraisemblance d'un élément de preuve valant 6 est estimée à 0,002 (Figure VII.10.).

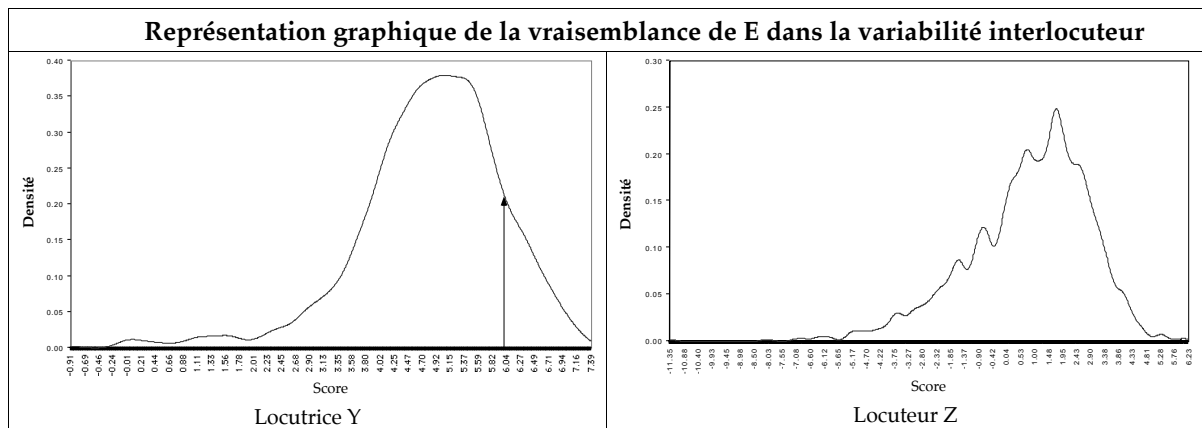


Figure VII.10. Calcul de la vraisemblance de E valant 6, dans le cas où l'hypothèse H_2 est vérifiée

7.3.5.3. Rapport de vraisemblance de l'élément de preuve E

Le rapport de vraisemblance est obtenu en divisant $p(E | H_1)$ par $p(E | H_2)$ ⁹², et est calculé de la manière suivante :

$$LR = \frac{\frac{1}{m} \sum_{i=1}^m K(E | a_i, \lambda)}{\frac{1}{n} \sum_{i=1}^n K(E | b_i, \lambda)} \quad (7.12)$$

⁹¹ *supra*: 3.5.5.1.5. Formalisation

⁹² *supra*: 3.5.5.1.5. Formalisation

Par exemple pour la locutrice Y, le rapport de vraisemblance des hypothèses H_1 et H_2 vaut 1,81 (0,40 / 0,22), pour un élément de preuve E valant 6 (Figure VII.11.).

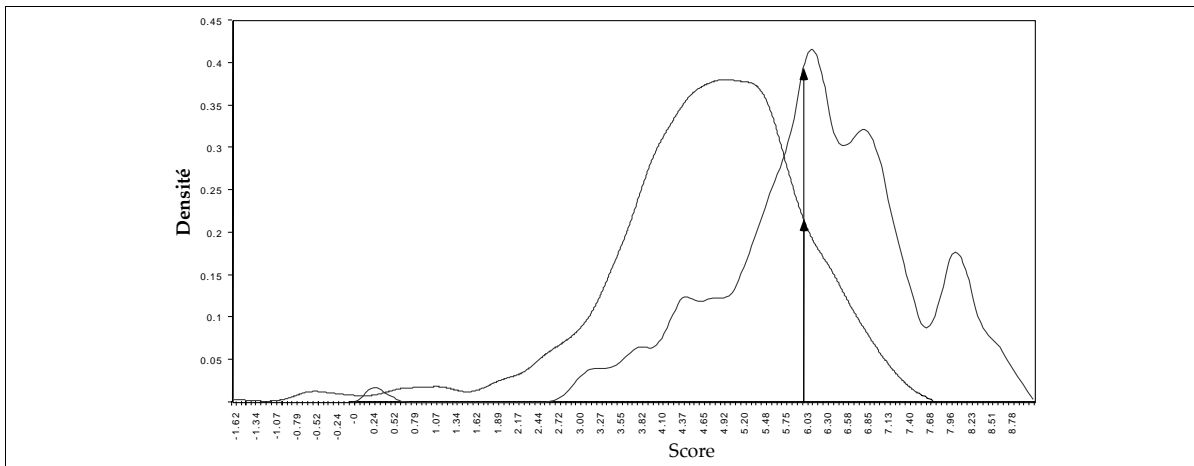


Figure VII.11. Représentation graphique du rapport de vraisemblance de H_1 et H_2 pour la locutrice Y

Pour le locuteur Z, le rapport de vraisemblance des hypothèses H_1 et H_2 est estimé à 75 (0,15 / 0,002), pour un élément de preuve E valant 6 (Figure VII.12.). Cet exemple met aussi en évidence l'importance du dénominateur dans le calcul des rapports de vraisemblance.

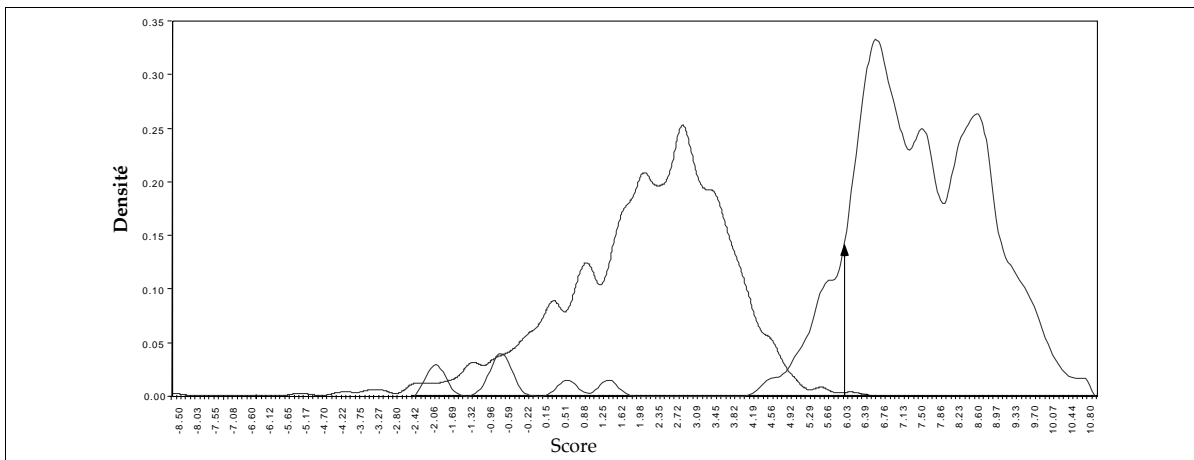


Figure VII.12. Représentation graphique du rapport de vraisemblance de H_1 et H_2 pour le locuteur Z

7.4. Expériences

7.4.1. Principe

Les expériences réalisées au chapitre VIII servent à tester le système de reconnaissance automatique de locuteurs dans différentes conditions rencontrées en criminalistique. Le principe d'évaluation de la méthode consiste à estimer et à comparer la distribution des rapports de vraisemblance qui peuvent être obtenus à partir de l'élément de preuve e , d'une part lorsque

l'hypothèse H_1 est vérifiée, c'est-à-dire lorsque la source du modèle et celle de l'enregistrement de test est unique, et d'autre part lorsque l'hypothèse H_2 est vérifiée, c'est-à-dire lorsque la source du modèle et celle de l'enregistrement de test est différente.

7.4.2. Présentation des résultats

Le choix s'est porté sur un mode de présentation proposé par EVETT ET BUCKLETON dans le domaine de l'interprétation de l'analyse génétique forensique ; ces deux auteurs ont choisi de nommer ce mode de présentation « *Tippet plot* », en référence aux concepts d'intravariabilité (*within-source comparison*) et d'intervariabilité (*between-source comparison*) définis par TIPPET ET AL.⁹³ [TIPPET ET AL., 1968 ; EVETT ET BUCKLETON, 1996 ; Evett, 2000].

L'axe des abscisses est gradué en termes de valeurs croissantes de rapports de vraisemblance (LR). L'axe des ordonnées indique la probabilité que le résultat du test excède une valeur de LR donnée. Chaque représentation graphique comporte deux courbes, la première rend compte de l'évolution des rapports de vraisemblance estimés lorsque l'hypothèse H_1 est vérifiée, alors que la seconde rend compte de la distribution des rapports de vraisemblance estimés lorsque l'hypothèse H_2 est vérifiée.

7.4.3. Exemple

L'exemple suivant illustre ce mode de présentation. Huit locuteurs ont enregistré chacun six modèles de leur voix et un enregistrement de test, considéré comme l'indice. Pour chaque locuteur, la méthode de reconnaissance calcule six indices de proximité en comparant l'enregistrement de test aux six modèles de voix, ce qui donne globalement 48 valeurs (6 indices de proximité * 8 locuteurs). Ces valeurs sont les éléments de preuve E lorsque l'hypothèse H_1 est vérifiée. Pour chacun de ces 48 éléments de preuve E un rapport de vraisemblance est calculé⁹⁴. L'évolution de ces 48 rapports de vraisemblance est illustrée par la courbe grise de la figure VII.13.

Ensuite, pour chaque locuteur, la méthode de reconnaissance compare l'enregistrement de test aux 1000 modèles de voix des locuteurs de la base de données représentant la population potentielle, ce qui donne globalement 8000 valeurs (1000 indices de proximité * 8 locuteurs). Ces valeurs sont les éléments de preuve E lorsque l'hypothèse H_2 est vérifiée. Pour chacun de ces 8000 éléments de preuve E un rapport de vraisemblance est calculé⁹⁵. L'évolution de ces 8000 rapports de vraisemblance est illustrée par la courbe noire du graphique VII.13.

Ce mode de présentation illustre de manière simultanée les performances du système lorsque l'une ou l'autre des deux hypothèses alternatives H_1 ou H_2 est vérifiée.

⁹³ *supra* : 3.6.4.1. Critères de sélection des bases de données

⁹⁴ *supra* : 7.3. Méthode de calcul du rapport de vraisemblance

⁹⁵ *supra* : 7.3. Méthode de calcul du rapport de vraisemblance

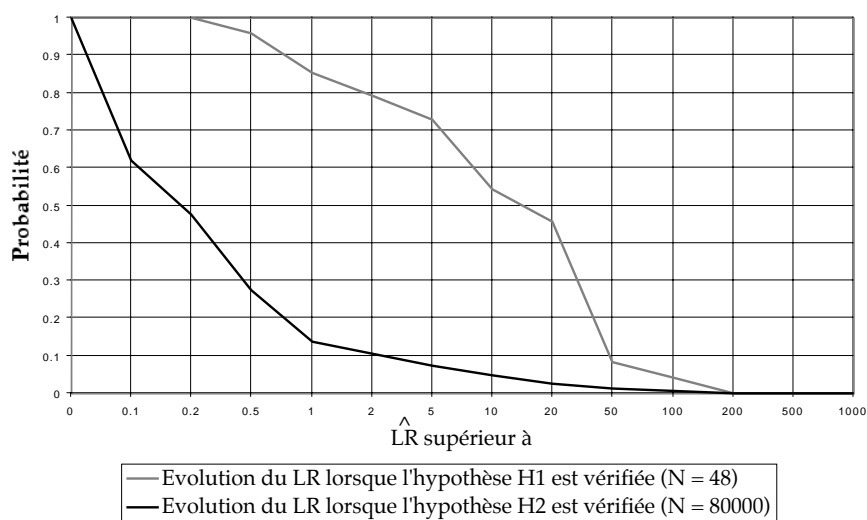


Figure VII.13. Représentation graphique d'un résultat d'expérience sous forme de « *Tippet plot* »

7.5. Conclusion

Ce chapitre a principalement permis de décrire la structure et le fonctionnement du système de reconnaissance développé, qui est basé sur quatre modules logiciels : un module de suppression des silences « SILREM », un module d'extraction des caractéristiques « PLP », un module de modélisation « GMM » et un module de comparaison « GMM-evaluate ». Le calcul des rapports de vraisemblance repose quant à lui sur une estimation des variabilités intralocuteur et interlocuteur par *kernel density estimation* et les résultats sont présentés graphiquement sous forme de « *Tippet plot* ».

VIII. ÉVALUATION DU SYSTEME

8.1. Introduction

L'évaluation du système débute par la sélection et la constitution de bases de données d'enregistrements de parole. Il est important que la qualité de ces enregistrements soit comparable à celle qui peut être atteinte lors de l'enregistrement d'un message anonyme ou d'une écoute téléphonique. Les différentes conditions définies pour les enregistrements de comparaison et les enregistrements utilisés pour le calcul des modèles doivent servir à circonscrire au mieux la procédure nécessaire à l'obtention d'une évaluation réaliste de la variabilité intralocuteur. Quant aux différentes conditions présentes dans les enregistrements de test, les indices dans un cas réel, doivent permettre de cerner les limites d'application du système dans un cadre forensique.

L'évaluation du système se poursuit par l'estimation des limites théoriques du système de reconnaissance développé dans le chapitre VII et se termine par une évaluation des performances du système dans différentes conditions qui peuvent être rencontrées dans le cadre d'une application forensique. Cette dernière partie sert aussi à illustrer et à démontrer la capacité de la méthode à évaluer de manière satisfaisante tout système automatique de reconnaissance de locuteurs à vocation forensique.

8.2. Enregistrement et sélection de bases de données

Cette étape consiste à enregistrer ou à sélectionner les deux bases de données, la première servant à estimer la variabilité interlocuteur à l'intérieur de la population des locuteurs qui sont potentiellement à l'origine de l'enregistrement considéré comme indice, la seconde servant à l'estimation de la variabilité intralocuteur de la ou des personne(s) suspectée(s) d'être la source de l'indice. Elle sert aussi à constituer un ensemble d'enregistrements de test, de manière à simuler des indices matériels qui peuvent être rencontrés en cas d'abus de téléphone ou de mesure de surveillance.

8.2.1. Détermination de la langue parlée

Comme l'indépendance des méthodes automatiques de reconnaissance de locuteurs par rapport à la langue parlée n'est pas démontrée, la procédure d'évaluation développée dans cette recherche est réalisée dans une seule langue, le français.

8.2.2. Estimation de la variabilité intralocuteur

Une base de données, baptisée « Polyphone IPSC », a été intégralement enregistrée dans le cadre de cette recherche. Le rôle des personnes mises en cause a été tenu par 32 personnes, huit paires de femmes (A à H) et huit paires d'hommes (I à P), qui habitent la Suisse Romande et s'expriment en français (Tableau VIII.1. et Tableau VIII.2). Les personnes formant chaque paire ont

un lien de parenté entre elles et ont été sélectionnées sur la base d'une proximité auditive subjective de leur voix : en effet, leurs proches déclarent les confondre régulièrement au téléphone.

Paire	Locutrice	Langue maternelle	Âge	Partenaire	Lien de parenté	Appréciation subjective de la proximité auditive
A	00	allemande	24	01	Fille	Plus grande proximité en allemand qu'en français
	01	allemande	54	00	Mère	
B	04	française	32	06	Fille	Grande proximité au téléphone
	06	française	59	04	Mère	
C	05	française	26	49	Sœur	Grande proximité au téléphone
	49	française	27	05	Sœur	
D	07	française	31	08	Sœur	Grande proximité au téléphone
	08	française	33	07	Sœur	
E	09	française	64	33	Fille	Proximité moyenne au téléphone
	33	française	32	09	Mère	
E	32	française	13	44	Sœur jumelle	Grande proximité au téléphone
	44	française	13	32	Sœur jumelle	
G	54	française	52	55	Mère	Proximité moyenne au téléphone
	55	française	26	54	Fille	
H	58	française	54	59	Fille	Proximité moyenne au téléphone
	59	française	25	58	Mère	

Tableau VIII.1. Les locutrices de la base de données « Polyphone IPSC »

Paire	Locuteur	Langue maternelle	Âge	Partenaire	Lien de parenté	Appréciation subjective de la proximité auditive
I	10	française	61	56	Père	Grande proximité au téléphone
	56	française	29	10	Fils	
J	11	française	31	20	Fils	Grande proximité au téléphone
	20	française	62	11	Père	
K	12	française	30	40	Fils	Proximité moyenne au téléphone
	40	française	55	12	Père	
L	13	française	36	41	Frère	Grande proximité au téléphone
	41	française	40	13	Frère	
M	14	française	33	15	Frère	Grande proximité au téléphone
	15	française	35	14	Frère	
N	16	française	23	17	Fils	Grande proximité au téléphone
	17	française	56	16	Père	
O	18	française	33	19	Frère jumeau	Grande proximité au téléphone
	19	française	33	18	Frère jumeau	
P	22	française	77	39	Père	Proximité moyenne au téléphone
	39	française	49	22	Fils	

Tableau VIII.2. Les locuteurs de la base de données « Polyphone IPSC »

8.2.2.1. Acquisition des signaux de parole

8.2.2.1.1. Système d'enregistrement

En Suisse, les transmissions entre les centraux téléphoniques sont presque exclusivement numériques. Par contre, la police et les fournisseurs d'accès aux télécommunications réalisent encore souvent leurs enregistrements depuis le réseau téléphonique, à l'aide d'équipements analogiques de mauvaise qualité ⁹⁶.

L'enregistrement numérique des conversations téléphoniques est techniquement possible en Suisse, mais cette évolution technologique dépend surtout d'une volonté politique de fournir les moyens nécessaires à l'obtention d'une qualité d'enregistrement adéquate. Pour cette raison, l'acquisition des signaux téléphoniques utilisés dans cette recherche a été réalisée directement depuis le réseau téléphonique numérique sur une plate-forme informatique, par l'intermédiaire d'une ligne de type RNIS et d'une carte de conversion.

Les données ont été acquises sur une station de travail UNIX de marque et modèle *Sun*[®] *ULTRASPARC*[™], reliée au réseau téléphonique par l'intermédiaire d'une carte d'acquisition de type *SunISDN-BRI/SBI*[™], qui permet un enregistrement des données au format G.711 ⁹⁷. Le pilotage de la partie matérielle a été assuré par le logiciel *Sunlink*[™] ISDN 1.0, développé par *Sun Microsystems*[®] et livré avec la carte d'acquisition. Le même type de matériel a été utilisé par *Swisscom*[®] pour l'enregistrement de la base de données nommée « Polyphone Suisse Romande ».

8.2.2.1.2. Justification de la procédure d'enregistrement par téléphone

Les enregistrements réalisés dans le cadre pénal sont, en grande majorité, enregistrés par l'intermédiaire du réseau téléphonique ⁹⁸. Dans ce cas, les enregistrements servant de comparaison devraient aussi être enregistrés par téléphone, selon l'étude de HUNT [HUNT, 1983]. Cette procédure de collecte des enregistrements de comparaison par le réseau téléphonique a aussi l'avantage de pouvoir être appliquée directement dans la réalité policière. Il est en effet simple pour un fonctionnaire de police de collecter les enregistrements de comparaison en demandant à une personne mise en cause d'effectuer une série d'enregistrements par téléphone depuis une pièce calme de taille moyenne. Cette manière de procéder aboutit aussi à une meilleure standardisation de la qualité des enregistrements de comparaison que des prises de son effectuées directement par les différents services de police, dans des conditions non contrôlées avec un matériel disparate.

8.2.2.1.3. Conversion et segmentation

Pour l'analyse, les données ont été converties du format non linéaire de 8 bits - 8 kHz (G.711) au format linéaire de 16 bits - 8 kHz, avec le logiciel *Audiotool*[™] livré avec le système d'exploitation *Solaris*[™] 2.5. Ce logiciel a aussi été utilisé pour la segmentation et l'édition des fichiers qu'il permet de réaliser avec une précision de l'ordre d'un dixième de seconde.

⁹⁶ *supra*: 2.3.6. Influence du système d'enregistrement

⁹⁷ *supra*: 2.3.3.2. Réseau téléphonique public commuté (RTPC)

⁹⁸ *supra*: 2.1. Introduction

8.2.2.2. Composition de la base de données « Polyphone IPSC »

8.2.2.2.1. Enregistrements réalisés pour la modélisation

Pour modéliser la voix des 16 locutrices et 16 locuteurs, chaque participant a réalisé six sessions d'enregistrement de données d'entraînement, comparables aux sessions de la base de données « Polyphone Suisse Romande », sur le plan du contenu et de la qualité technique. Une première session, intitulée « Session Polyphone cellulaire », a été enregistrée avec un téléphone cellulaire, (GSM)⁹⁹, mis à la disposition des participants. Les autres sessions ont été intitulées « Session Polyphone 1 à 5 ». En principe, chaque personne les a enregistrées avec son téléphone privé ou professionnel, avec fil ou sans fil¹⁰⁰. Signe des temps, certains participants ont dérogé à cette règle en utilisant une ou plusieurs fois leur propre téléphone cellulaire (Annexe VI.2.a.). Ces six enregistrements ont été effectués sur une période de un à trois mois selon les participants (Annexe VI.1.a.). Une septième session d'enregistrement, intitulée « Session Comparaison » a été utilisée pour la réalisation d'un modèle : elle est composée des deux premières minutes de la session d'enregistrement de comparaison.

8.2.2.2.2. Enregistrements de comparaison

Ce type d'enregistrements vise à estimer la variabilité intralocuteur des 16 locutrices et 16 locuteurs dans différents styles d'élocution. Pour y parvenir, chaque personne a été amenée à commenter la même session d'une cinquantaine de diapositives. L'enregistrement dure de 5 à 15 minutes selon les locuteurs. Chaque participant a enregistré cette session avec son téléphone privé ou professionnel, avec fil ou sans fil (Annexe VI.2.b.).

La majorité des diapositives est constituée d'images à commenter de manière spontanée, alors qu'une plus petite partie est constituée de quelques questions, de lecture avec un crayon dans la bouche, de dialogues à simuler et de messages anonymes à proférer, sans déguisement de la voix ou avec un crayon dans la bouche. Pour chaque personne, l'enregistrement a eu lieu le jour « J », sans préparation, de manière à laisser une large place à la spontanéité, voire à la surprise. Mis à part les deux premières minutes de cette session, qui ont été utilisées pour constituer le septième modèle de la voix de chacun des 32 participants, l'enregistrement a été segmenté manuellement en énoncés de 1 à 30 s, de 12 à 43 énoncés selon la loquacité des personnes (Annexe VI.3a., b., c. et d.).

8.2.3. Estimation de la variabilité interlocuteur

Comme la constitution d'une base de données de grande taille est très onéreuse et demande beaucoup de temps, un moyen consiste à rechercher et à acquérir une base de données existante et répondant aux critères mis en évidence dans l'indice auprès d'un organisme comme ELRA (*European Language Resources Association*) en Europe, ou LDC (*Linguistic Data Consortium*) aux États-Unis. Ce choix souffre cependant d'une restriction : la possibilité de procéder à des sessions d'enregistrement dont la qualité technique est strictement comparable à celles présentes dans la

⁹⁹ *supra* : 2.3.3.3. Réseau téléphonique cellulaire

¹⁰⁰ *supra* : 2.3.3.2. Réseau téléphonique public commuté (RTPC)

base de données choisie, avec la personne mise en cause. Comme toutes les personnes citées *supra* proviennent de Suisse Romande et s'expriment en français, l'estimation de la variabilité interlocuteur a été mesurée sur une base de données qui modélise cette population, la base de données « Polyphone Suisse Romande ».

8.2.3.1. Sélection de la base de données « Polyphone Suisse Romande »

8.2.3.1.1. Critère de choix de la base de données

La base de données « Polyphone Suisse Romande » a été gracieusement mise à disposition par le laboratoire *R & D Digital Signal Processing* de l'entreprise *Swisscom*[®]. Cette base de données comprend les énoncés de 2500 locutrices et 2500 locuteurs ; la base de données Polyphone est disponible chez ELRA (*European Language Resources Association*) en plusieurs autres langues, à savoir l'allemand, l'anglais, le hollandais et depuis peu le suisse allemand.

Pour la constituer, les locuteurs ont été choisis au hasard en Suisse Romande par l'entreprise *Swisscom*[®]. Chaque personne enregistrée a effectué une session d'enregistrement depuis son téléphone privé, professionnel ou cellulaire. Les sessions ont été enregistrées par l'intermédiaire d'une ligne téléphonique numérique RNIS au format G.711 ; elles ont une durée de 90 à 150 secondes, selon les personnes, et sont principalement constituées de texte lu et de parole spontanée, sous forme d'une demande de renseignement téléphonique.

8.2.3.1.2. Dimensionnement de la base de données

La population de la Suisse Romande comprend moins de 2 millions de personnes. Pour des raisons de taille des données, de temps de modélisation et de comparaison, deux sous-ensembles de cette base de données ont été sélectionnés pour représenter la population de Suisse Romande : 1000 locutrices (n° 0001 à 1000) et 1000 locuteurs (n° 4001 à 5000). Ces deux sous-ensembles permettent de constituer non seulement une bonne image de la variabilité des locutrices et des locuteurs de cette région, mais aussi de la variabilité du réseau téléphonique public commuté et cellulaire de Suisse Romande. L'utilisation du téléphone cellulaire était cependant moins répandue en 1995 qu'elle ne l'est aujourd'hui.

8.2.3.1.3. Énoncés considérés pour la modélisation et pour l'évaluation des limites théoriques du système

Pour chaque locuteur de cette base de données, l'énoncé contenant la demande de renseignement téléphonique, qui dure une dizaine de secondes, a été séparé des autres énoncés. Il a été utilisé comme enregistrement de test, pour évaluer les limites théoriques du système de reconnaissance¹⁰¹. Les 80 à 140 secondes restantes ont été utilisées comme données d'entraînement, pour calculer les modèles statistiques de la voix des 1000 locutrices et 1000 locuteurs sélectionnés.

¹⁰¹ *infra* : 8.4. Limites théoriques du système

8.2.4. Constitution d'enregistrements de test

Les 32 personnes sélectionnées pour jouer le rôle des personnes mises en cause ont aussi contribué à constituer un ensemble d'enregistrements de test, simulant les indices qui peuvent être rencontrés en cas d'abus de téléphone ou de mesure de surveillance.

8.2.4.1. Simulation d'abus de téléphone

Pour simuler le type d'enregistrement recueilli en cas d'abus de téléphone, les participants ont premièrement effectué deux messages anonymes, l'un sans déguisement de la voix et l'autre avec un déguisement de leur choix. Dans les deux cas, le contenu et la longueur du message étaient laissés à la liberté de chacun (Annexe VI.3.e., f., g. et h.).

8.2.4.2. Simulation de mesures de surveillance

Pour simuler le type d'enregistrement recueilli dans le cadre d'une mesure de surveillance, les demandes de renseignement enregistrées dans le cadre des six sessions « Polyphone IPSC » ont été retirées de ces sessions pour servir d'enregistrement de test ; elles sont intitulées « Test cellulaire » et « Test 1 » à « Test 5 » (Annexe VI.3.e., f., g. et h.).

Le « Test 1 » a aussi servi à la fabrication de plusieurs enregistrements secondaires. Il a été enregistré sur le système d'enregistrement des conversations téléphoniques de la Police Cantonale de Neuchâtel. Cet enregistreur, de marque et modèle *Atis*[®] VCG-600 analogique, fabriqué en Suisse par *Atis-Uher* SA à Fontaines (NE), permet 24 heures d'enregistrement de 40 pistes simultanées sur une seule bande magnétique de type VHS. La qualité de l'enregistrement résultant est suffisante pour conserver une intelligibilité acceptable, mais insuffisante pour sauvegarder intégralement la qualité du signal provenant du réseau téléphonique. Finalement le « Test 1 » a été utilisé pour fabriquer une série d'enregistrements de test bruités, par addition d'un bruit de fond enregistré lors d'un apéritif dans une salle contenant une centaine de personnes. Huit enregistrements ont été produits avec un rapport signal sur bruit de 0, 3, 6, 9, 12, 18, 24 et 30 dB (Annexe VI.3.e., f., g. et h.).

En principe, les participants ont réalisé ces enregistrements de test avec leur téléphone privé ou professionnel, sauf le test réalisé avec le téléphone cellulaire mis à leur disposition. Quelques personnes ont aussi utilisé leur propre téléphone cellulaire pour l'enregistrement d'une session de test (Annexe VI.2.c).

8.3. Procédure d'évaluation du système

La première phase de l'évaluation consiste à évaluer les limites théoriques du système, sur la base des deux bases de données « Polyphone Suisse Romande » et « Polyphone IPSC ».

Dans une deuxième phase, l'évaluation du système consiste à mettre en cause chacune des 32 personnes, en comparant ses enregistrements de test avec les modèles de sa propre voix et avec les modèles des voix des personnes de la population potentielle. Cette configuration de test permet d'évaluer les rapports de vraisemblance qui peuvent être dégagés lorsque la personne mise en cause est la source réelle de l'indice matériel.

Dans une troisième phase, l'évaluation consiste à mettre en cause, dans chaque paire, la seconde personne de la paire, en comparant les enregistrements de test de la première personne avec les modèles de la voix de la seconde et avec les modèles des voix des personnes de la population potentielle. Cette configuration de test permet d'évaluer les rapports de vraisemblance qui peuvent être dégagés lorsque la personne mise en cause n'est pas la source réelle de l'indice matériel, mais une personne dont la voix est auditivement proche de celle de l'auteur.

Dans une dernière phase, l'évaluation consiste à mettre en cause chacune des 32 personnes, en comparant ses enregistrements de test avec les modèles de sa propre voix et avec les modèles de la voix de la seconde personne de la paire. Cette configuration de test permet d'évaluer les rapports de vraisemblance qui peuvent être dégagés lorsque l'hypothèse alternative proposée, par exemple par la défense, indique que la source est une personne dont la voix est auditivement proche.

8.4. Limites théoriques du système

8.4.1. Évaluation sur la base de données « Polyphone Suisse Romande »

8.4.1.1. Procédure

Cette évaluation est réalisée sur une sélection de 500 locutrices (n° 0001 à 0500) et de 500 locuteurs (n° 4001 à 4500). L'enregistrement de test est comparé au modèle de la voix des 500 personnes de la population potentielle de même genre (n° 0001 à 0500 ou n° 4001 à 4500) et les indices de proximité sont classés par ordre décroissant. Ce test est une classification en ensemble fermé¹⁰², puisqu'à chaque fois le modèle de l'enregistrement testé se trouve dans la base de données.

Le calcul d'un rapport de vraisemblance n'est pas possible dans cette situation ; en effet il n'existe qu'un seul enregistrement pour chaque personne de la base de données « Polyphone Suisse Romande », ce qui empêche toute évaluation de l'intravariabilité.

8.4.1.2. Résultats

Pour 485 des 500 locutrices, un enregistrement de test existe. A l'issue de la classification en ensemble fermé, la locutrice correspondant au plus grand indice de proximité est la vraie locutrice à 466 reprises. A 19 reprises ce n'était pas le cas, ce qui correspond à un taux de fausse identification de 3,9 % pour les locutrices (Tableau VIII.3.).

Pour 472 des 500 locuteurs, il existe un enregistrement de test. A l'issue de la classification en ensemble fermé, le locuteur correspondant au plus grand indice de proximité est le vrai locuteur à 452 reprises. A 20 reprises ce n'était pas le cas, ce qui correspond à un taux de fausse identification de 4,2 % pour les locuteurs (Tableau VIII.3.).

¹⁰² *supra* : 3.5.2.1.1. Classification en ensemble fermé (*closed set*)

Rang	N	1 ^e	2 ^e	3 ^e	4 ^e	5 ^e	6 ^e	7 ^e	8 ^e	9 ^e	10 ^e et au delà	FA (%)
Locutrices	485	466	8	3	1	1	1	1	1	1	2	3,9 %
Locuteurs	472	452	15	3	1	0	0	1	0	0	2	4,2 %

Tableau VIII.3. Résultats de la classification en ensemble fermé dans une sélection de 500 locutrices et 500 locuteurs de la base de données « Polyphone Suisse Romande »

8.4.2. Évaluation sur la base de données « Polyphone IPSC »

8.4.2.1. Procédure

Cette évaluation a été réalisée en utilisant les enregistrements de test « Test 1 » à « Test 5 » et « Test cellulaire » des 32 participants. Pour chaque personne, ces enregistrements de test sont comparés, d'une part aux modèles provenant de la même session d'enregistrement que le test et d'autre part aux modèles des voix des personnes de la base de données « Polyphone Suisse Romande ». Pour les participantes, les enregistrements de test sont comparés aux modèles des locutrices (n° 0001 à 1000) et pour les locuteurs ils sont comparés aux modèles des locuteurs (n° 4001 à 5000). Ce test est aussi une classification en ensemble fermé ¹⁰³.

8.4.2.2. Résultats

A l'issue de la classification en ensemble fermé, le plus grand indice de proximité a été mis en évidence à 90 reprises pour les locutrices et à 91 reprises pour les locuteurs, alors que l'enregistrement de test provenait de la même source que le modèle. Sur la base de 96 tests, ces résultats représentent des taux de fausse identification de 6,3 % pour les femmes et 5,2 % pour les hommes (Tableau VIII.4.).

	Nombre de tests	1 ^e rang	FA (%)
Locutrices	96	90	6,3 %
Locuteurs	96	91	5,2 %

Tableau VIII.4. Résultats de la classification en ensemble fermé dans la base de données « Polyphone IPSC »

8.4.3. Discussion des résultats

Les taux de fausse identification obtenus dans ces deux évaluations permettent d'évaluer les limites théoriques du système sur la base d'enregistrements téléphoniques contemporains. Ils sont à rapprocher des résultats de FLOCH, qui obtient un taux de fausse identification de 4,4 % avec une méthode indépendante du texte sur l'entier de la base de données « TIMIT », composée d'appels téléphoniques non bruités de 192 locutrices et 438 locuteurs [FLOCH ET AL. 1994 ; FISHER ET AL., 1986]. Ces résultats servent à vérifier le fonctionnement du système, mais ne reflètent en aucun cas les performances de la méthode dans des conditions forensiques.

Les taux d'erreur plus faibles obtenus sur la base de données « Polyphone Suisse Romande » que sur la base de données « Polyphone IPSC » s'expliquent de deux manières. Premièrement les

¹⁰³ *supra* : 3.5.2.1.1. Classification en ensemble fermé (*closed set*)

enregistrements de test sont comparés à 500 modèles dans le premier cas et à mille modèles dans le second ; deuxièmement les personnes formant la base de données « Polyphone Suisse Romande » sont *a priori* non liées entre elles alors que les personnes formant la base de données « Polyphone IPSC » sont *a priori* liées deux par deux.

Les valeurs d'erreur obtenues dans la première évaluation peuvent être considérées comme précises puisqu'elles résultent de 458'009 tests mettant en jeu 957 personnes. L'imprécision qui frappe les valeurs mises en évidence dans la seconde évaluation est par contre beaucoup plus importante, puisqu'elles proviennent de 96'000 tests mettant en jeu seulement 32 locuteurs *a priori* liés deux par deux. Par exemple les six erreurs sont concentrées sur la locutrice L05 dans l'évaluation des locutrices et pour les locuteurs, quatre des cinq erreurs sont concentrées sur le locuteur L39.

Cette constatation met aussi en évidence l'hétérogénéité des performances de reconnaissance parmi les locuteurs d'une population et laisse à penser que les participants L05 et L39 sont des personnes spécialement difficiles à reconnaître alors que les autres personnes de la base de données « Polyphone IPSC » sont des locuteurs dont les tests fournissent des résultats conformes aux attentes. Ce phénomène est connu dans le domaine de la reconnaissance automatique de locuteurs ; DODDINGTON, par exemple, désigne ces deux types de locuteurs différents par les termes de « chèvres » et de « moutons » [DODDINGTON *ET AL.*, 1998].

Finalement, cette expérience montre les limites de la classification comme méthode d'inférence de l'identité en sciences forensiques. En effet, l'information selon laquelle l'auteur se situe au 2^{ème} rang des 500 locuteurs est intéressante en sciences forensiques, alors que seule la configuration dans laquelle l'auteur est positionné au premier rang est pertinente pour un contrôle d'accès.

8.5. Évaluation de l'influence du temps séparant l'enregistrement de l'indice et celui du modèle

8.5.1. Procédure

La voix se modifie au cours du temps et la durée entre l'enregistrement de l'indice et du modèle est susceptible d'altérer les performances du système de reconnaissance automatique de locuteurs. L'influence de ce paramètre est évaluée à l'aide des enregistrements de test « Test 1 » à « Test 5 » des 32 participants à la base de données « Polyphone IPSC », enregistrés sur une période de deux mois pour les locutrices et de trois pour les locuteurs.

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de test « Test 1 » à « Test 5 » avec les cinq modèles de sa propre voix « Session Polyphone 1 » à « Session Polyphone 5 ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison des mêmes enregistrements de test « Test 1 » à « Test 5 » avec les modèles de la voix des 1000 locutrices ou des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement de test. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.5.2. Résultats

Cette procédure de test amène deux types de configurations : dans la première, l'enregistrement de test est enregistré avant le modèle alors que dans la seconde, il est enregistré après le modèle.

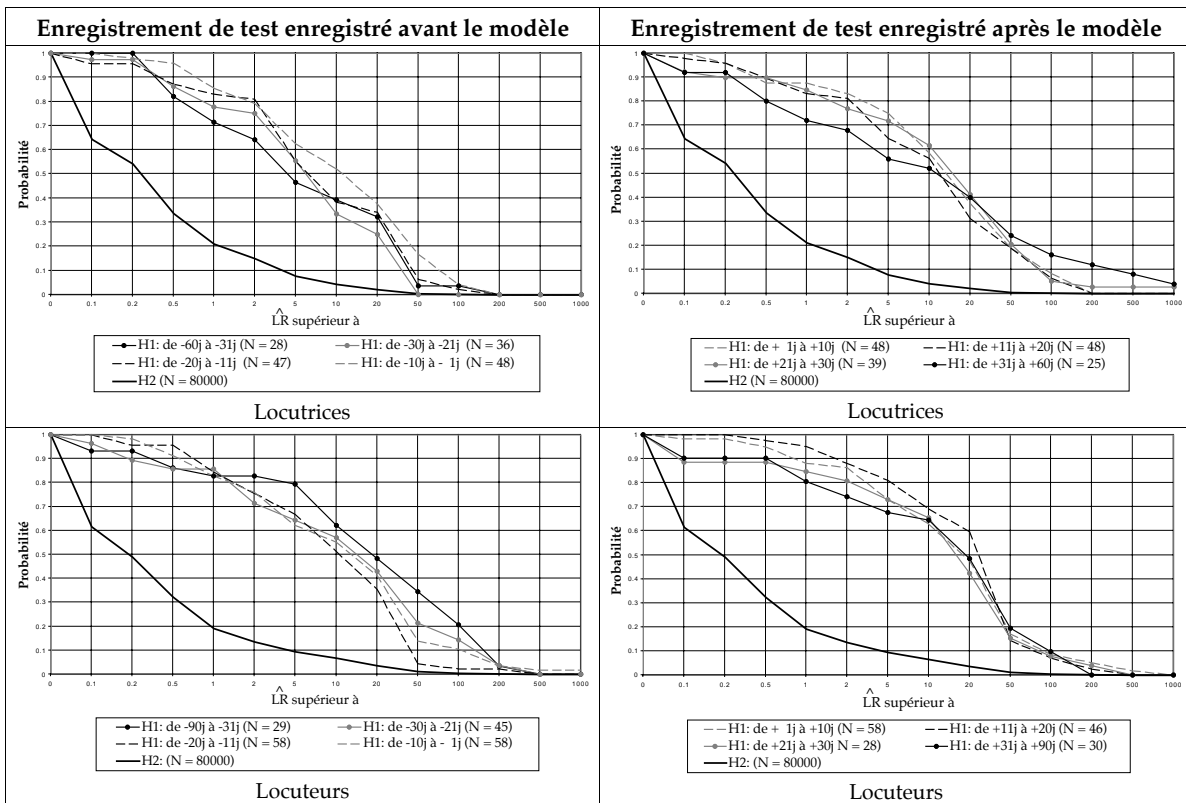


Figure VIII.1. Résultat de l'évaluation globale des rapports de vraisemblance en fonction du temps séparant l'enregistrement de test de l'enregistrement du modèle

8.5.3. Discussion des résultats

Dans les applications commerciales, le test est généralement enregistré après le modèle. En sciences forensiques par contre, la configuration où l'enregistrement de l'indice est réalisé avant l'enregistrement du modèle est la règle. La configuration opposée n'est pas impossible mais improbable ; elle peut exister dans le cas où la personne mise en cause continue à être écoutée ou à effectuer des appels anonymes après que la police a recueilli un enregistrement de comparaison de sa voix.

Les résultats montrent que l'influence du temps qui sépare l'enregistrement du modèle de celui de l'indice existe, mais elle ne semble pas prépondérante pour les durées étudiées, au maximum deux mois pour les locutrices et trois pour les locuteurs (Figure VIII.1). De plus, pour les locuteurs, aucune tendance claire ne peut être déterminée lorsque la trace est enregistrée avant le modèle. Finalement, les performances sont légèrement meilleures lorsque le test est enregistré après le modèle, ce qui représente un handicap dans la configuration de la plupart des cas forensiques.

8.6. Évaluation de l'influence de la quantité et de la qualité des données

8.6.1. Influence du type d'élocution lors de l'enregistrement des modèles

8.6.1.1. Procédure

La qualité du modèle réalisé à partir de la voix de la personne suspectée est susceptible de varier en fonction du type d'élocution adopté, parole lue ou parole spontanée. L'influence de ce paramètre est évaluée à l'aide des modèles «Session Comparaison » et «Session Polyphone 1 » des 32 participants à la base de données « Polyphone IPSC ». Ces deux sessions ont été enregistrées avec le même téléphone à une demi-heure d'intervalle. Le modèle provenant de la « Session Comparaison » est composé d'une quantité plus importante de parole spontanée que de parole lue, alors que le modèle issu de la « Session Polyphone 1 » est constitué en majorité de parole lue.

Comme il s'agit d'une comparaison directe des performances lorsque deux modèles différents sont utilisés, seule la situation où l'hypothèse H_1 est vérifiée a été prise en compte dans cette expérience. Les éléments de preuve E sont le résultat de la comparaison des enregistrements de test « Test 2 » à « Test 5 » de chacun des 32 locuteurs de la base de données « Polyphone IPSC », soit avec les modèles «Session Comparaison », soit avec les modèles «Session Polyphone 1 ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement de test. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.6.1.2. Résultats

Pour chaque locutrice et chaque locuteur, deux séries de quatre éléments de preuve, et de là deux séries de quatre rapports de vraisemblance, ont été calculées. Les résultats sont d'abord présentés de manière globale, pour les locutrices et pour les locuteurs. Ensuite, ils sont présentés de manière individuelle ; pour chaque locutrice et chaque locuteur deux rapports de vraisemblance moyens, exprimés en termes de logarithmes en base 10, sont calculés à partir des deux séries de rapports de vraisemblance, de manière à évaluer l'influence de la qualité des données qui constituent le modèle.

8.6.1.2.1. Type d'élocution lors l'enregistrement du modèle : évaluation globale

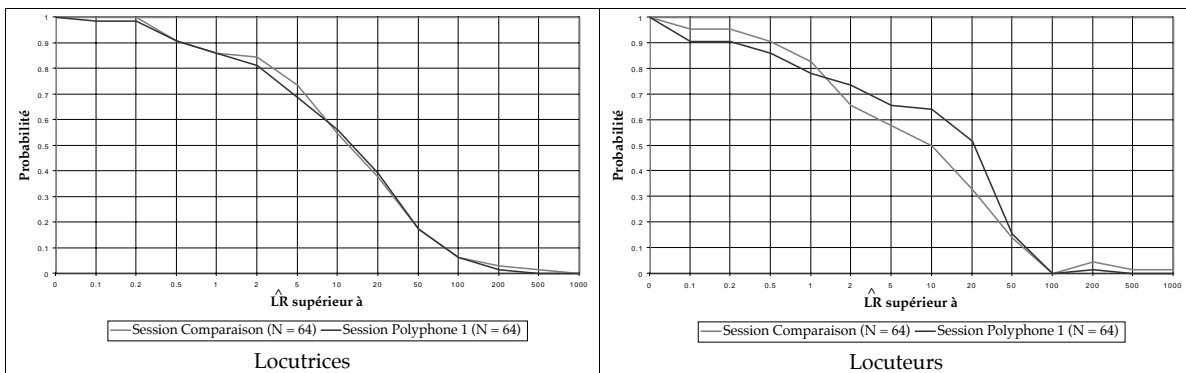


Figure VIII.2. Résultat de l'évaluation globale des rapports de vraisemblance, en fonction du type d'élocution adopté lors de l'enregistrement des modèles

8.6.1.2.2. Type d'élocution lors l'enregistrement du modèle : évaluation individuelle

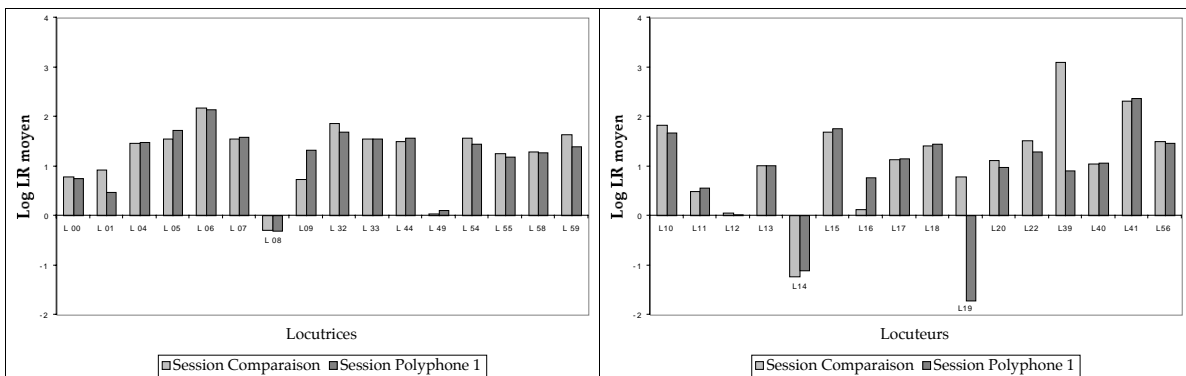


Figure VIII.3. Résultat de l'évaluation individuelle des rapports de vraisemblance moyens, en fonction du type d'élocution adopté dans l'enregistrement utilisé pour la modélisation de la voix

8.6.1.3. Discussion des résultats

L'influence du contenu et du type d'élocution adopté pour l'enregistrement du modèle n'est pas importante, ce qui confirme le caractère indépendant du texte de la méthode GMM. En effet, les résultats obtenus à partir d'un modèle formé de parole spontanée sont très proches de ceux tirés d'un modèle composé d'une majeure partie de parole lue, ce qui peut être observé tant de manière globale (Figure VIII.2) que de manière individuelle (Figure VIII.3). Ce résultat permet de choisir indifféremment une base de données composée de parole lue ou de parole spontanée pour l'évaluation de la variabilité interlocuteur.

Par contre, les résultats présentés de manière individuelle (Figure VIII.3) mettent encore une fois en évidence la différence qui existe entre la majorité des locuteurs, dont les tests fournissent des résultats conformes, et une minorité, les locuteurs L08, L12, L14, L49 et partiellement L19, dont les résultats sont contraires aux attentes (Figure VIII.2.).

8.6.2. Influence de la quantité de parole dans les enregistrements de comparaison

8.6.2.1. Procédure

La qualité de l'évaluation de l'intravariabilité d'un locuteur est susceptible d'être influencée par la durée des enregistrements de comparaison utilisés à cet effet. L'influence de ce paramètre est évaluée à l'aide des enregistrements de comparaison nommés « Parole spontanée » des 32 participants à la base de données « Polyphone IPSC ». Ces enregistrements ont été séparés en deux groupes : le premier contient les énoncés de parole d'une durée de 0 à 4 s et le second les énoncés de parole de plus de 4 s.

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de comparaison nommés « Parole spontanée » avec six modèles de sa propre voix : « Session Polyphone Cellulaire » et « Session Polyphone 1 » à « Session Polyphone 5 ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de comparaison nommés « Parole spontanée » avec les modèles de la voix des 1000 locutrices et des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement de comparaison. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de comparaison.

8.6.2.2. Résultats

Les personnes qui ont toujours utilisé le même téléphone sont évaluées indépendamment de celles qui ont utilisé des téléphones ou des lignes de téléphone différents pour l'enregistrement des modèles.

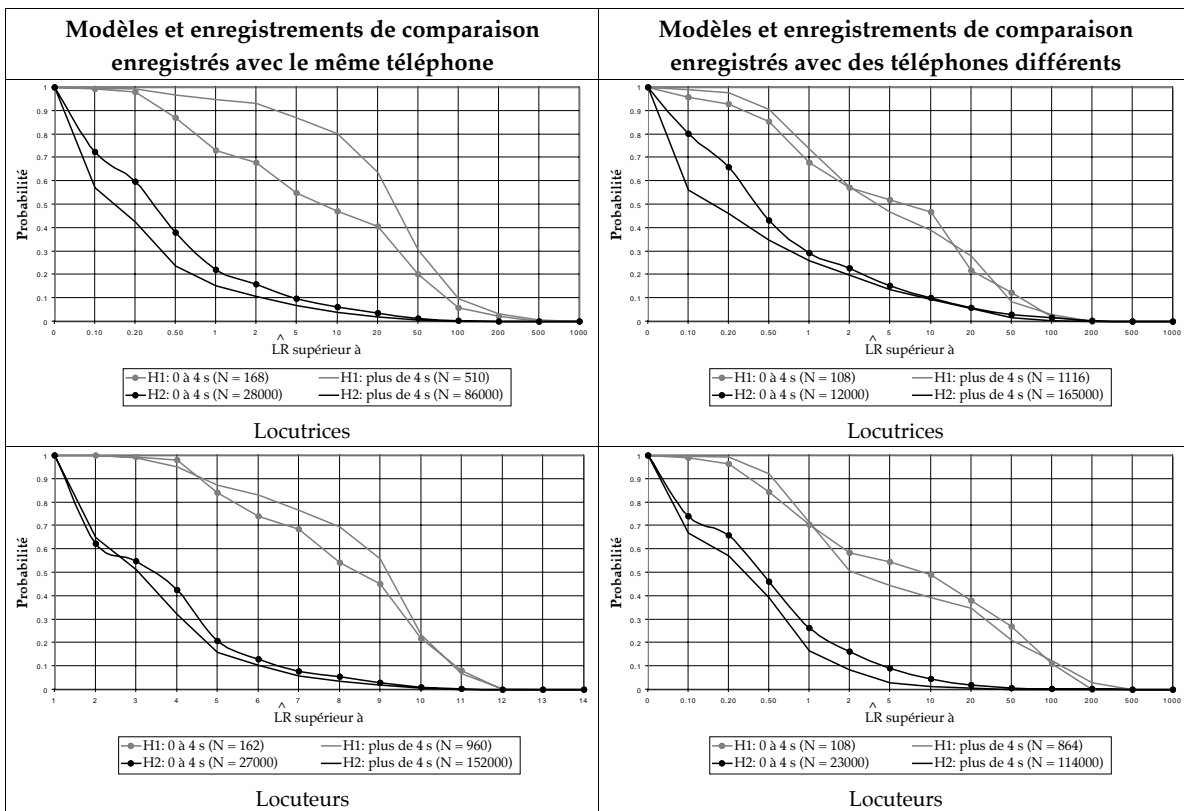


Figure VIII.4. Résultat de l'évaluation globale des rapports de vraisemblance, en fonction de la durée de l'enregistrement de comparaison

8.6.2.3. Discussion des résultats

Il est normal de constater qu'une diminution de la quantité de parole dans les enregistrements de comparaison altère les performances du système de reconnaissance automatique de locuteurs. Cependant, l'influence de la quantité de parole qui compose les enregistrements de comparaison est plus importante dans le cas où l'enregistrement de tous les modèles provient du même téléphone que lorsqu'ils proviennent de téléphones différents. L'introduction de la variabilité concernant le canal de transmission conduit à des performances globales inférieures pour les locutrices et locuteurs qui ont utilisé plusieurs téléphones pour l'enregistrement des modèles. Dans le même temps, les rapports de vraisemblance obtenus lorsque l'hypothèse H_1 est vérifiée diminuent et les rapports de vraisemblance obtenus lorsque l'hypothèse H_2 est vérifiée augmentent (Figure VIII.4.).

Dans le domaine forensique, ce résultat indique que les personnes mises en cause dont la voix est modélisée pour les besoins de l'enquête doivent utiliser des téléphones différents pour

l'enregistrement des modèles et les enregistrements de comparaison, de manière à ne pas diminuer artificiellement la variabilité intralocuteur par l'élimination du facteur lié au canal de transmission.

Dans le cas où le téléphone utilisé pour l'enregistrement de l'indice est connu et accessible, il est alors possible d'utiliser ce même et unique téléphone pour effectuer toutes les séances d'enregistrement de toutes les personnes mises en cause. Cette situation est idéale, car elle permet une amélioration potentielle du résultat en éliminant le facteur de variabilité lié au téléphone et à la ligne téléphonique.

8.6.3. Influence du type d'élocution dans les enregistrements de comparaison

8.6.3.1. Procédure

La qualité de l'évaluation de l'intravariabilité d'un locuteur est susceptible d'être influencée par le type d'élocution dans les enregistrements de comparaison utilisés à cet effet. L'influence de ce paramètre est évaluée à l'aide de deux types d'enregistrement de comparaison, les enregistrements nommés « Simulation de dialogues », dans lequel le locuteur joue des dialogues, et les enregistrements nommés « Lecture déguisée », dans lequel le locuteur lit un texte avec un crayon dans la bouche. Chaque type d'enregistrement a été séparé en deux groupes : le premier contient les énoncés de parole d'une durée de 0 à 4 s ou de 0 à 8 s et le second, les énoncés de parole de plus de 4 s ou de plus de 8 s.

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de comparaison nommés « Simulation de dialogues » et « Lecture déguisée » avec six modèles de sa propre voix nommés « Session Polyphone Cellulaire » et « Session Polyphone 1 » à « Session Polyphone 5 ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de comparaison nommés « Simulation de dialogues » et « Lecture déguisée » avec les modèles de voix « Session Polyphone » des 1000 locutrices et des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement de comparaison. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de comparaison.

8.6.3.2. Résultats

Les personnes qui ont toujours utilisé le même téléphone sont évaluées indépendamment de celles qui ont utilisé des téléphones ou des lignes de téléphone différents pour l'enregistrement des modèles.

8.6.3.2.1. Simulation de dialogues lors des enregistrements de comparaison

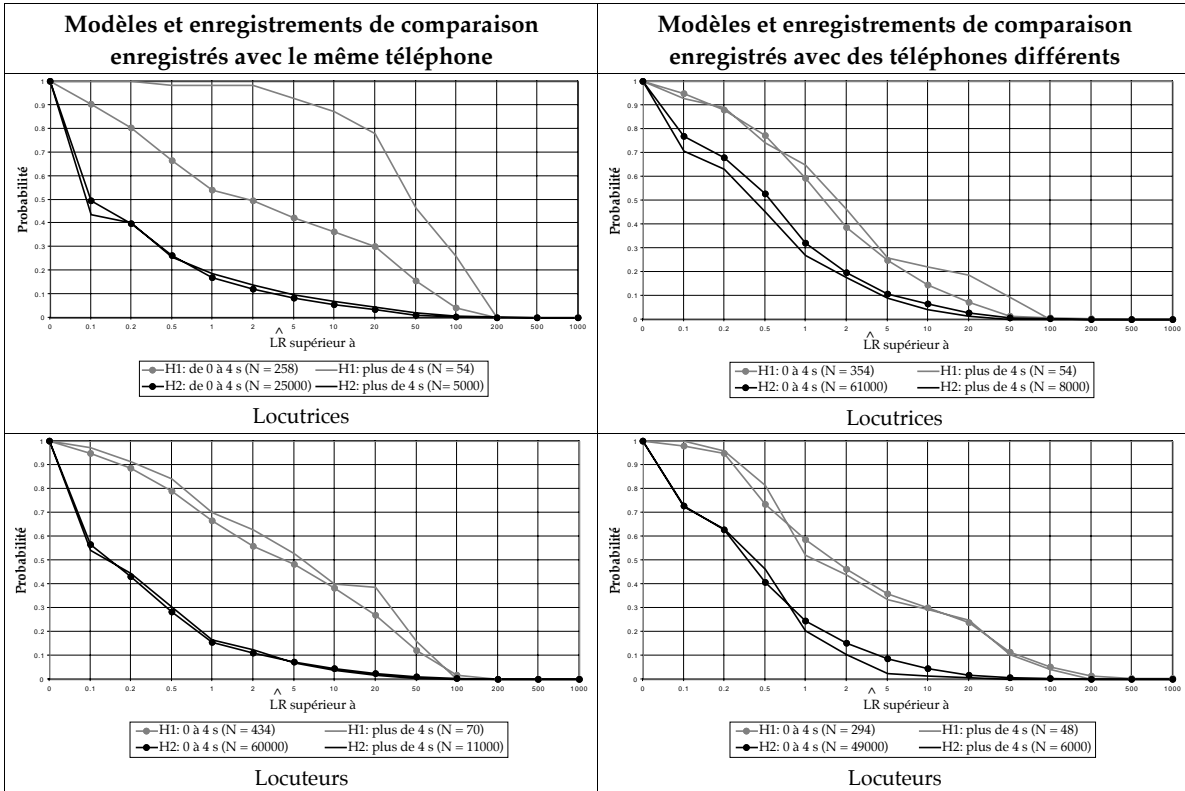
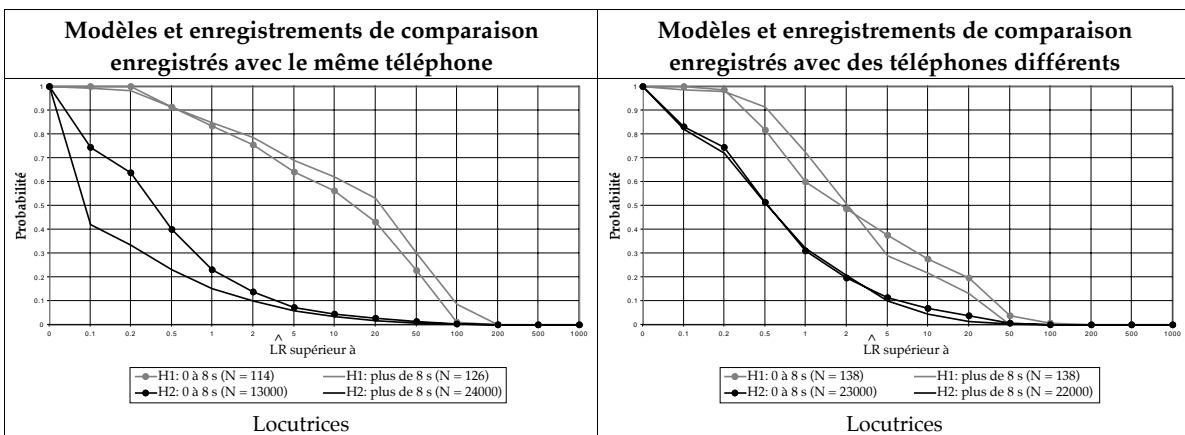


Figure VIII.5. Résultat de l'évaluation globale des rapports de vraisemblance lorsque les enregistrements de comparaison sont composés de dialogues simulés

8.6.3.2.2. Lecture déguisée lors des enregistrements de comparaison



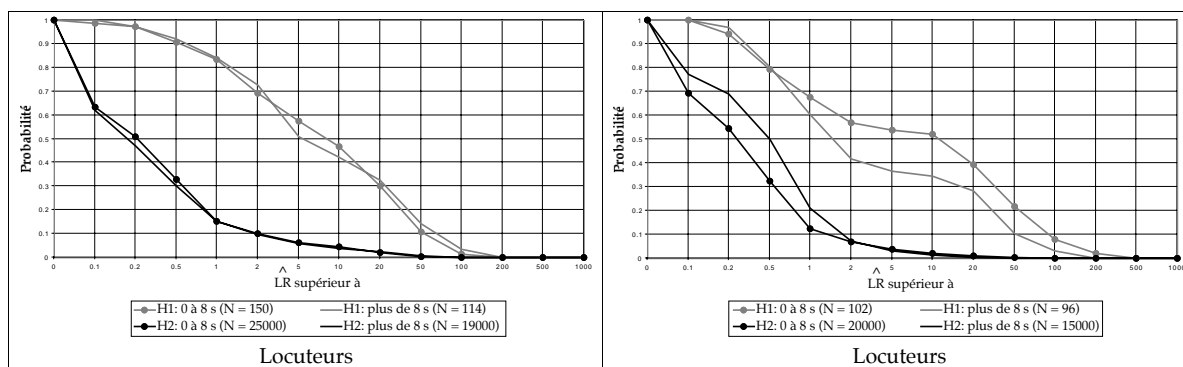


Figure VIII.6. Résultat de l'évaluation globale des rapports de vraisemblance lorsque les enregistrements de comparaison sont composés de lecture avec la voix déguisée

8.6.3.3. Discussion des résultats

L'influence de la qualité des données qui constituent les enregistrements de comparaison est importante. En effet, lorsque le style de l'élocution est particulier, que ce soit des dialogues simulés (Figure VIII.5.) ou de la lecture déguisée avec un crayon dans la bouche (Figure VIII.6.), les résultats sont inférieurs à ceux obtenus avec des enregistrements de parole spontanée (Figure VIII.4.). Par contre l'influence de la durée de ces enregistrements est peu importante, particulièrement lorsque les différents enregistrements ont été réalisés depuis plusieurs téléphones.

Dans le domaine forensique, ce résultat indique que la session d'enregistrement de comparaison doit proposer des exercices favorisant plusieurs styles d'élocution pour modéliser correctement la variabilité intralocuteur ; le choix devrait se baser sur le ou les styles d'élocution présents dans l'indice. Une session de commentaire de diapositives comparable à celle constituée pour l'enregistrement de la « Session Comparaison » de la base de données « Polyphone IPSC » est un moyen d'y parvenir. Sans préparation, cet exercice de lecture et de description d'images nécessite toute l'attention du locuteur et contribue à obtenir une spontanéité acceptable ; en effet, un contrôle conscient de l'élocution est difficile durant ce type d'exercice, car il nuit à la fluidité du discours et se remarque rapidement. L'exercice de simulation de dialogues est particulièrement concluant pour les personnes qui maîtrisent parfaitement la lecture, car elles se prennent facilement à ce jeu. Pour les personnes qui ne maîtrisent pas parfaitement la lecture, la description d'images reste le meilleur moyen d'obtenir une élocution fluide et spontanée.

8.7. Évaluation de l'influence d'un déguisement de la voix

8.7.1. Déguisement de la voix dans les enregistrements de comparaison

8.7.1.1. Procédure

La qualité de l'évaluation de l'intravariabilité d'un locuteur est susceptible d'être influencée par la présence d'un déguisement de la voix dans les enregistrements de comparaison utilisés à cet

effet. L'influence de ce paramètre est évaluée à l'aide d'enregistrements de comparaison nommés « Simulation de messages anonymes », composés de messages anonymes simulés et prononcés avec une voix normale ou déguisée par la présence d'un crayon dans la bouche lors de l'élocution.

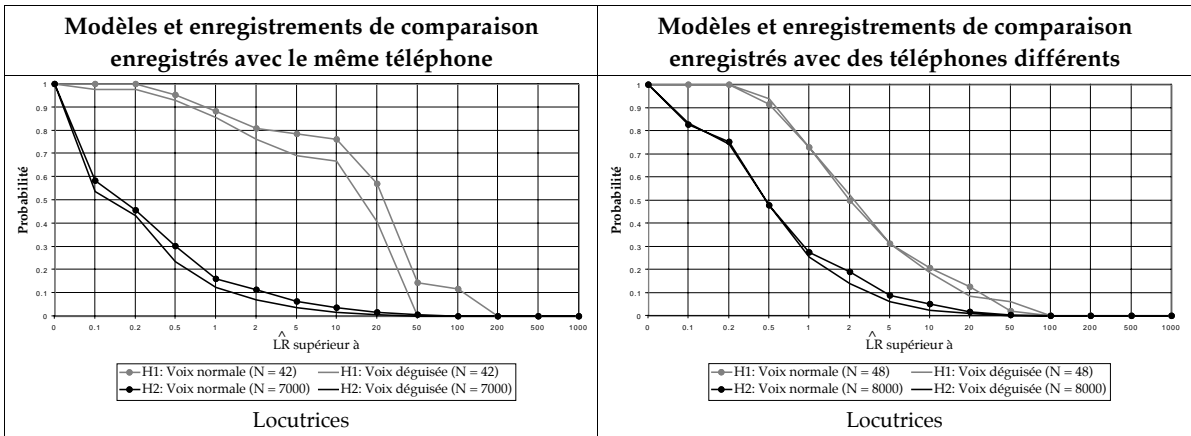
Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de comparaison nommés « Simulation de messages anonymes » avec les six modèles de sa propre voix « Session Polyphone Cellulaire » et « Session Polyphone 1 » à « Session Polyphone 5 ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de comparaison nommés « Simulation de messages anonymes » avec les modèles de la voix des 1000 locutrices et des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.7.1.2. Résultats

Les personnes qui ont toujours utilisé le même téléphone sont évaluées indépendamment de celles qui ont utilisé des téléphones ou des lignes de téléphone différents pour l'enregistrement des modèles.



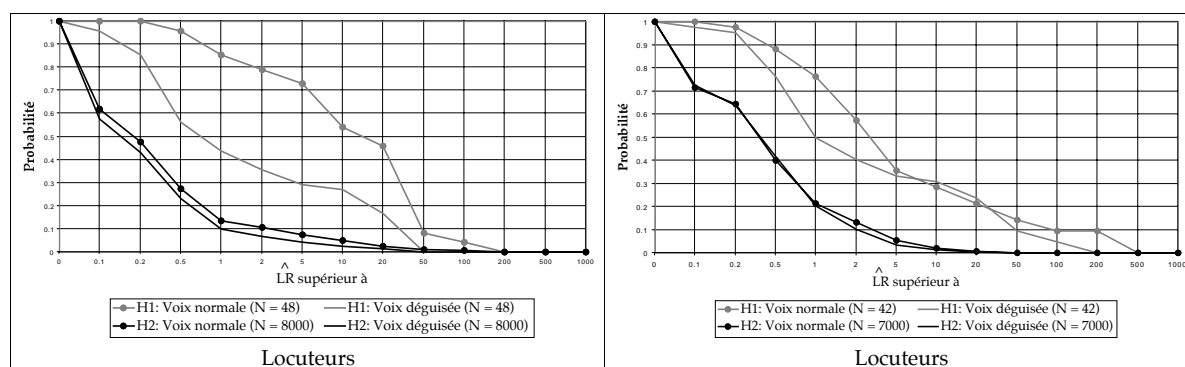


Figure VIII.7. Résultat de l'évaluation globale des rapports de vraisemblance lorsqu'il y a absence ou présence d'un déguisement de la voix dans l'enregistrement de comparaison

8.7.1.3. Discussion des résultats

La présence d'un déguisement de la voix dans l'enregistrement de comparaison altère les résultats par rapport à l'élocution normale. Curieusement, cette altération est moins importante pour les locutrices que pour les locuteurs. Une des raisons réside dans le fait que, dans la session de comparaison, le déguisement consistait, pour tous les locuteurs, à s'exprimer avec un crayon dans la bouche (Figure VIII.7.). Certaines personnes ont minimisé l'effet de cet élément gênant en le plaçant au coin de la bouche, par commodité. Dans les enregistrements de test, le déguisement était laissé à la liberté de chacun, ce qui s'est soldé par une dégradation des résultats beaucoup plus importante (Figure VIII.8.).

D'un point de vue forensique, ce premier résultat indique que le système automatique de reconnaissance de locuteurs est sensible à tout élément modifiant la morphologie et la physiologie du tractus vocal. Ce type d'exercice ne devrait pas prendre place dans l'enregistrement de comparaison, à moins de l'improbable démonstration que l'indice a été réalisé dans des conditions similaires.

8.7.2. Déguisement de la voix dans les enregistrements de test

8.7.2.1. Procédure

La présence d'un déguisement de la voix dans les enregistrements de test est susceptible d'altérer les performances du système de reconnaissance automatique de locuteurs. L'influence de ce paramètre est évaluée à l'aide d'enregistrements de test composés de messages anonymes simulés « Messages anonymes » et prononcés avec une voix normale ou déguisée. Comme le choix du déguisement était laissé à la liberté de chaque locutrice et locuteur, la stratégie de déguisement a aussi été analysée.

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de test « Messages anonymes » avec les six modèles de sa propre voix « Session Polyphone Cellulaire » et « Session Polyphone 1 » à « Session Polyphone 5 ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de comparaison nommés « Messages anonymes » avec les modèles de la voix des 1000 locutrices et des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.7.2.2. Résultat

Les personnes qui ont toujours utilisé le même téléphone sont évaluées indépendamment de celles qui ont utilisé des téléphones ou des lignes de téléphone différents pour l'enregistrement des modèles.

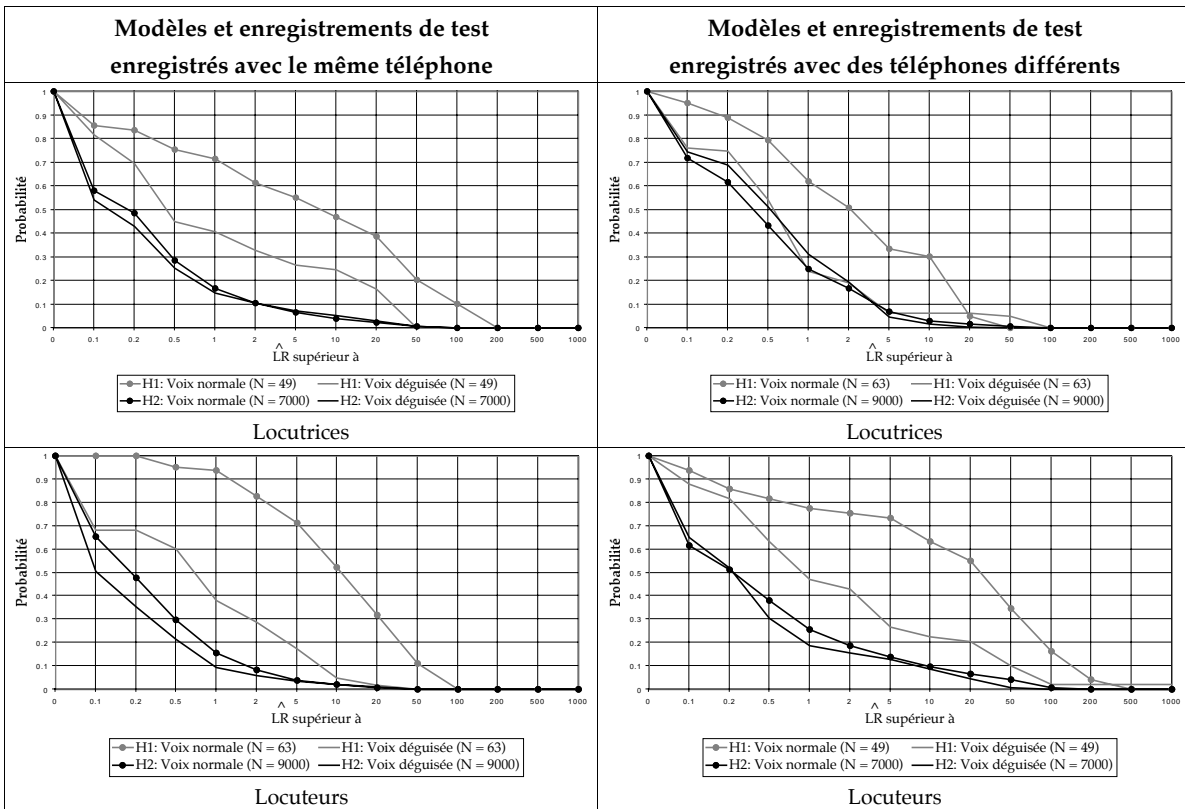


Figure VIII.8. Résultat de l'évaluation globale des rapports de vraisemblance lorsqu'il y a absence ou présence d'un déguisement de la voix dans l'indice

8.7.2.3. Stratégie de déguisement

Type de déguisement	Locutrices	Locuteurs
Nez bouché	5 (31,25 %)	2 (12,5 %)
Accent étranger ou imitation	0	6 (37,5 %)
Objet gênant l'élocution	5 (31,25 %)	1 (6,25 %)
Élocution lente	1 (6,25 %)	1 (6,25 %)
Élévation de F_0	1 (6,25 %)	2
Abaissement de F_0	1 (6,25 %)	0
Mouchoir devant le microphone	1 (6,25 %)	1 (6,25 %)
Colère	1 (6,25 %)	0
Élocution lente + objet gênant l'élocution	1 (6,25 %)	0
Abaissement de F_0 + voix rauque	0	2 (12,5 %)
Nez bouché + voix rauque	0	1 (6,25 %)

Tableau VIII.5. Type de déguisement adopté par les participants à la base de données « Polyphone IPSC »

8.7.2.4. Discussion des résultats

La présence d'un déguisement de la voix dans l'indice influence le résultat de manière prépondérante et confirme que le système de reconnaissance automatique de locuteurs n'est pas robuste à ce type de variabilité intralocuteur. Tous les types de déguisement choisis par les participants se sont révélés efficaces. En présence d'un déguisement, les chances d'obtenir un rapport de vraisemblance supérieur à 1 n'atteignent pas 50 %, alors que l'hypothèse H_1 est vérifiée (Figure VIII.8.). Par contre, la présence d'un déguisement n'a que peu d'influence sur les rapports de vraisemblance lorsque H_2 est vérifiée, contrairement à la variation du canal de transmission qui elle, a une grande influence sur les rapports de vraisemblance lorsque l'hypothèse H_2 est vérifiée (Figure VIII.8.). Lorsque les locutrices et les locuteurs ont utilisé des téléphones différents pour l'enregistrement des modèles et les enregistrements de test, l'altération des performances est si importante que les rapports de vraisemblance calculés sont très proches ou confondus, que l'hypothèse H_1 ou H_2 soit vérifiée. D'un point de vue forensique ce résultat indique que la méthode n'est pas utilisable lorsqu'il existe une suspicion de déguisement de la voix dans l'indice.

Le résultat de l'analyse des stratégies de déguisement utilisées par les participants à la base de données « Polyphone IPSC » rejoint les résultats de l'étude de MASTHOFF¹⁰⁴ sur plusieurs points [MASTHOFF, 1996]. Il a aussi été observé que, lorsque le contenu est court, une seule phrase, les locuteurs optent majoritairement pour la modification d'un seul paramètre (87,5 %). De même, environ 10 % des personnes utilisent un système de filtrage extrinsèque, un mouchoir devant le microphone dans la présente expérience, et une dizaine de stratégies différentes sont utilisées, de manière fort différente selon le sexe de la personne (Tableau VIII.5.).

¹⁰⁴ *supra* : 2.3.6.2. Enregistrement dans le cadre d'un abus de téléphone

Par contre d'autres points divergent ; par exemple la majorité des participants a opté pour une modification de l'articulation, notamment une modification du tractus vocal, alors que, dans l'étude de MASTHOFF, une majorité des personnes a procédé à une modification de la phonation. L'élévation et l'abaissement de la fréquence fondamentale est une stratégie adoptée par les femmes et les hommes, alors que MASTHOFF observe que les femmes privilégient l'abaissement de F_0 et les hommes son élévation (Tableau VIII.5.) [MASTHOFF, 1996].

8.8. Évaluation de l'influence du réseau, de la ligne et du téléphone

8.8.1. Influence du téléphone et de la ligne téléphonique utilisés pour l'enregistrement des modèles

8.8.1.1. Procédure

Le téléphone et la ligne téléphonique utilisés pour l'enregistrement des modèles sont susceptibles d'influencer la qualité de la modélisation de la voix de la personne suspectée, s'ils sont différents de ceux utilisés pour les enregistrements de comparaison. L'influence de ce paramètre est évaluée à l'aide des modèles « Session Polyphone 1 » à « Session Polyphone 5 ».

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des cinq modèles « Session Polyphone 1 » à « Session Polyphone 5 » avec les enregistrements de comparaison nommés « Parole spontanée ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de comparaison nommés « Parole spontanée » avec les modèles de la voix des 1000 locutrices et des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.8.1.2. Résultats

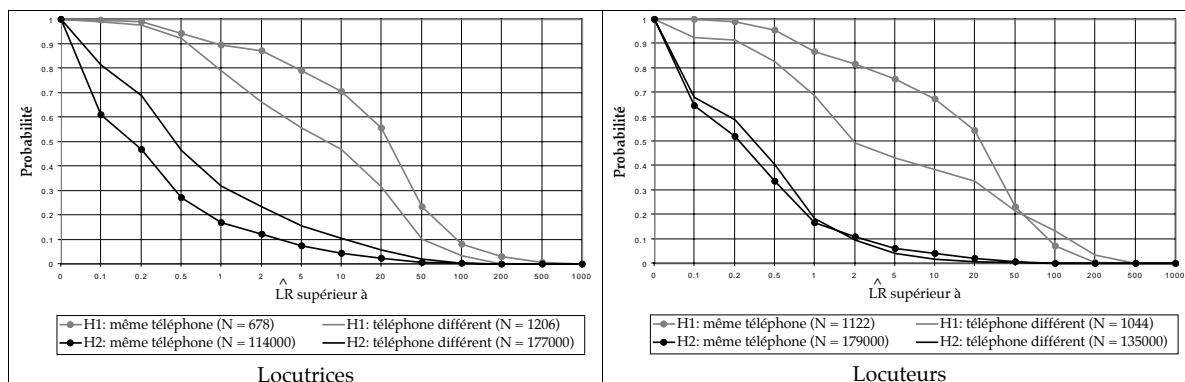


Figure VIII.9. Résultat de l'évaluation globale des rapports de vraisemblance en fonction du téléphone et de la ligne utilisés pour l'enregistrement des modèles

8.8.1.3. Discussion des résultats

Les résultats confirment les indications observées dans les expériences précédentes et reportées dans la littérature [PRZYBOCKI ET MARTIN, 1998]. L'utilisation d'un téléphone différent pour l'enregistrement du modèle et de la comparaison a une influence prépondérante sur le résultat. L'utilisation du même téléphone pour l'enregistrement du modèle et l'enregistrement de comparaison permet d'obtenir, dans 50 % des cas, un rapport de vraisemblance de l'ordre de 25 alors que l'on doit se contenter de rapports de vraisemblance de l'ordre de 7,5 pour les locutrices et de 2 pour les locuteurs, en cas d'utilisation de téléphones différents (Figure VIII.9.).

D'un point de vue forensique cette variabilité, introduite par la ligne téléphonique et le téléphone, indique qu'il est nécessaire de procéder à l'enregistrement des modèles à l'aide de plusieurs téléphones, différents de celui utilisé pour l'enregistrement de comparaison, sous peine de sous-évaluer artificiellement la variabilité intralocuteur.

8.8.2. Influence du téléphone et de la ligne téléphonique utilisés pour les enregistrements de test

8.8.2.1. Procédure

Le téléphone et la ligne téléphonique utilisés pour les enregistrements de test sont susceptibles d'influencer les performances de la reconnaissance, s'ils sont différents de ceux utilisés pour la modélisation. L'influence de ce paramètre est évalué à l'aide des enregistrements de test « Test 1 » à « Test 5 ».

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de test « Test 1 » à « Test 5 » avec les cinq modèles « Session Polyphone 1 » à « Session Polyphone 5 ». Les résultats provenant des comparaisons « Session Polyphone 1 » – « Test 1 », « Session Polyphone 2 » – « Test 2 » etc. n'ont pas été pris en compte car le modèle et l'indice proviennent dans ce cas de la même session d'enregistrement. Au total 20 rapports de vraisemblance sont calculés pour chaque locutrice et chaque locuteur.

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de test « Test 1 » à « Test 5 » avec les modèles de la voix des 1000 locutrices et des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.8.2.2. Résultats

Les résultats sont d'abord présentés de manière globale, afin d'illustrer les différences de performances entre les groupes qui ont toujours utilisé le même téléphone et les groupes qui ont utilisé des téléphones différents. Ensuite, les résultats sont présentés de manière individuelle ; pour chaque locutrice et chaque locuteur un rapport de vraisemblance moyen, exprimé en termes de logarithme en base 10, est calculé à partir des 20 rapports de vraisemblance calculés, de manière à illustrer les différences de performances entre les personnes.

8.8.2.2.1. Évaluation globale

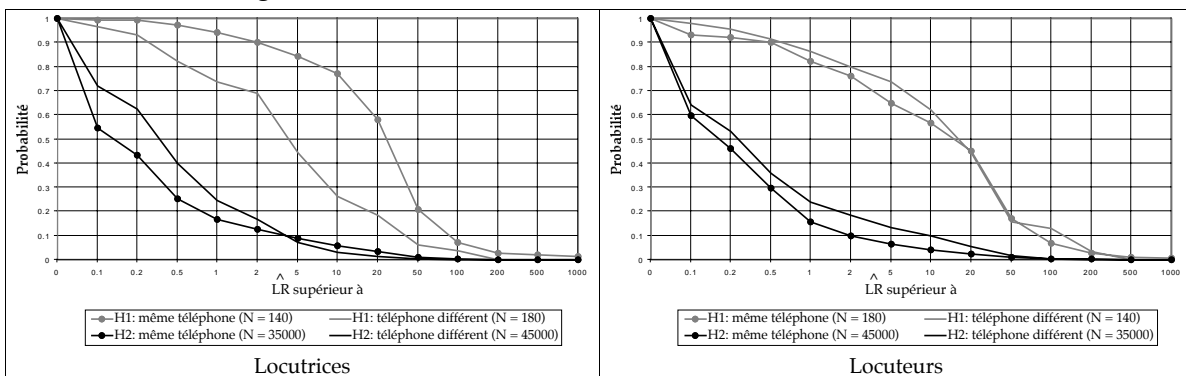


Figure VIII.10. Résultat de l'évaluation globale des rapports de vraisemblance en fonction du téléphone et de la ligne utilisés pour les enregistrements de test

8.8.2.2.2. Évaluation individuelle

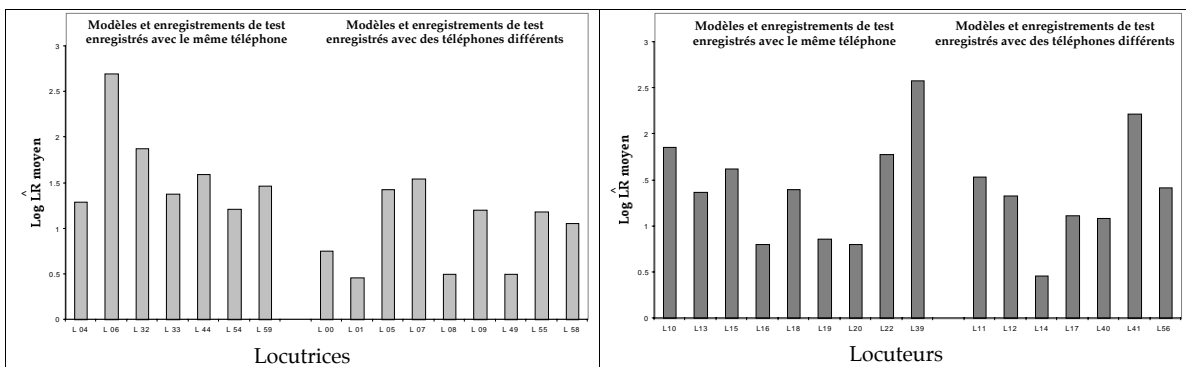


Figure VIII.11. Résultat de l'évaluation individuelle des rapports de vraisemblance moyens, en fonction du téléphone et de la ligne utilisés pour les enregistrements de test

8.8.2.3. Discussion des résultats

Dans l'évaluation globale, les résultats ne sont pas uniformes ; pour les locuteurs, ils montrent que l'utilisation de téléphones différents pour l'enregistrement du modèle et de l'indice n'a pas d'influence notable sur les performances. Par contre la méthode semble moins robuste pour les locutrices, lorsqu'elles utilisent des téléphones différents (Figure VIII.10). Comme la majorité des méthodes de reconnaissance automatique de locuteur, la méthode GMM a principalement été développée et testée avec des bases de données composées de voix d'hommes ; il n'est dès lors pas très surprenant que ses performances soient supérieures avec les voix d'hommes qu'avec les voix de femmes. Ce constat apparaît plusieurs fois dans les résultats.

L'évaluation individuelle met encore une fois en évidence les grandes disparités de performance du système automatique de reconnaissance de locuteurs entre les différentes personnes (Figure VIII.11). Cette grande variabilité est un défaut connu des méthodes de reconnaissance de locuteurs et, en l'absence d'améliorations permettant de diminuer ou d'éliminer ce phénomène, une évaluation extensive de l'intravariabilité de chaque personne soumise à une analyse reste une nécessité. L'utilisation d'un téléphone sans fil numérique de type *Digital Enhanced Cordless Telecommunication* (DECT) en lieu et place d'un téléphone filaire est sans influence notable.

8.8.3. Influence du réseau utilisé pour l'enregistrement des modèles

8.8.3.1. Procédure

Le fait que le modèle de la voix du locuteur soit enregistré à partir du réseau téléphonique public commuté (RTPC) ou du réseau téléphonique cellulaire (GSM) est susceptible d'influencer les performances du système automatique de reconnaissance de locuteurs, surtout si le réseau qui a servi à la production des enregistrements de test est différent de celui qui a servi à produire le modèle. L'influence de ce paramètre est évaluée à l'aide des deux modèles «Session Polyphone 1» et «Session Polyphone Cellulaire». Pour chaque personne de la base de données «Polyphone IPSC», les deux enregistrements ont été effectués le même jour, mais de manière indépendante, à environ une demi-heure d'intervalle.

8.8.3.2. Résultats

8.8.3.2.1. Modèle RTPC / GSM – Test RTPC

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données «Polyphone IPSC», de la comparaison des deux modèles «Session Polyphone 1» et «Session Polyphone Cellulaire» avec les enregistrements de comparaison nommés «Parole spontanée».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de comparaison «Parole spontanée» avec les modèles de la voix des 1000 locutrices et des 1000 locuteurs de la base de données «Polyphone Suisse Romande». Comme la base de données ne contient pas de session enregistrée avec un téléphone cellulaire, les enregistrements de comparaison n'ont pas pu être comparés à des modèles de la base de données enregistrés à partir d'un téléphone cellulaire.

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

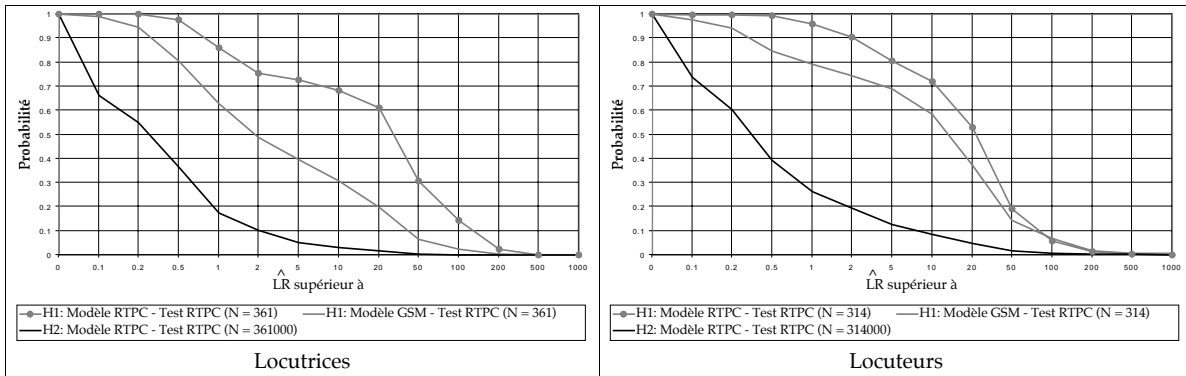


Figure VIII.12. Résultat de l'évaluation globale des rapports de vraisemblance lorsque le réseau téléphonique utilisé pour l'enregistrement des modèles est de type RTPC

8.8.3.2.2. Modèle GSM – Test RTPC /GSM

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison du modèle calculé à partir de la session « Session Polyphone Cellulaire » avec les enregistrements de test nommés « Test 1 » et « Test cellulaire ». Pour chaque locutrice et chaque locuteur de la base de données « Polyphone IPSC », ces deux enregistrements ont été effectués le même jour, mais de manière indépendante, à environ une demi-heure d'intervalle.

Comme la base de données « Polyphone Suisse Romande » ne contient pas de session enregistrée avec un téléphone cellulaire, les enregistrements de test n'ont pas pu être comparés à des modèles de la base de données enregistrés à partir d'un téléphone cellulaire.

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

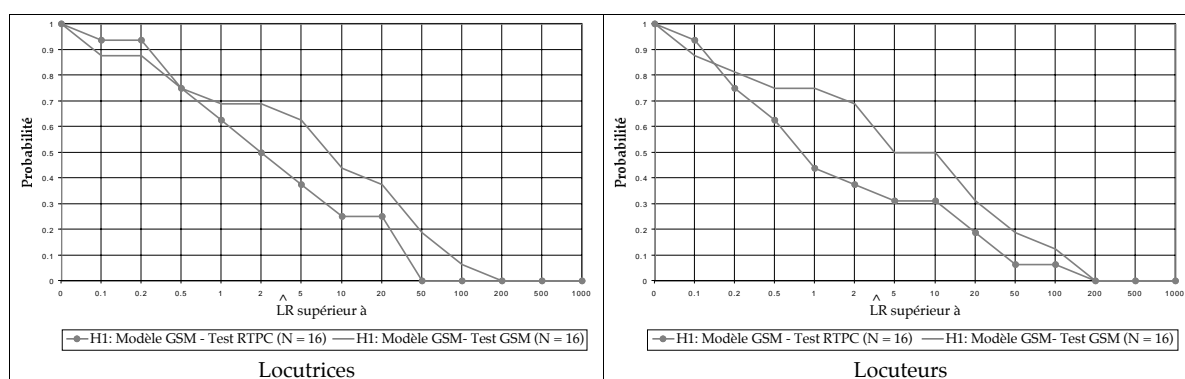


Figure VIII.13. Résultat de l'évaluation globale des rapports de vraisemblance lorsque le réseau téléphonique utilisé pour l'enregistrement des modèles est de type GSM

8.8.3.3. Discussion des résultats

L'utilisation d'un réseau cellulaire pour l'enregistrement du modèle à la place d'un réseau téléphonique commuté altère les performances du système de reconnaissance automatique de locuteurs, que les enregistrements de test proviennent du réseau téléphonique commuté ou cellulaire (Figure VIII.12 et VIII.13). Ce résultat n'est pas surprenant, car la méthode de codage de la parole utilisée dans le réseau cellulaire assure une fidélité au signal de base moindre que celle assurée par le système de codage de la parole dans le réseau commuté. En effet, dans le premier la parole est codée avec un débit de 16 Kbits s^{-1} , alors que dans le second ce débit est de 64 Kbits s^{-1} .

Du point de vue forensique, ce résultat indique qu'il est nécessaire de connaître le type de réseau par lequel a été transmis l'indice, de manière à réaliser l'enregistrement des modèles et les enregistrements de comparaison avec le même type de réseau téléphonique ; dans ce cas seulement la méthode de codage de la parole sera homogène dans tous les enregistrements.

8.8.4. Influence du réseau utilisé pour la production des enregistrements de test

8.8.4.1. Procédure

Le type de réseau téléphonique utilisé pour la production des enregistrements de test est susceptible d'influencer les performances du système de reconnaissance de locuteurs, surtout s'il est différent de celui utilisé pour les enregistrements servant à la modélisation. L'influence de ce paramètre est évaluée à l'aide des enregistrements de test « Test cellulaire » et « Test 1 ». Pour chaque locutrice et chaque locuteur de la base de données « Polyphone IPSC », ces deux enregistrements ont été effectués le même jour, mais de manière indépendante, à environ une demi-heure d'intervalle.

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de test « test cellulaire » et « test 1 » au modèle « Session Comparaison ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de test « test cellulaire » et « test 1 » avec les modèles

de la voix des 1000 locutrices et des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.8.4.2. Résultats

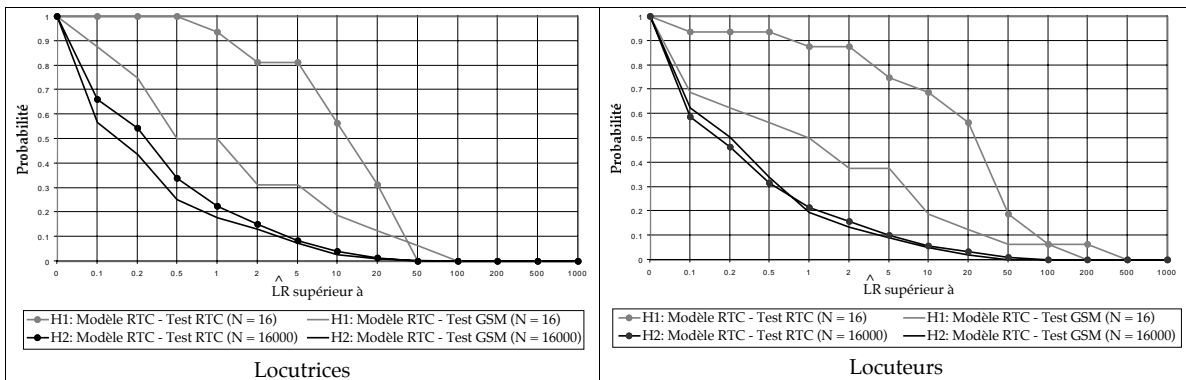


Figure VIII.14. Résultat de l'évaluation globale des rapports de vraisemblance en fonction du réseau téléphonique utilisé pour les enregistrements de test

8.8.4.3. Discussion des résultats

Les performances sont très nettement diminuées lorsque l'enregistrement du test provient du réseau cellulaire (GSM) et le modèle, du réseau téléphonique public commuté (RTPC) (Figure VIII.14.). Ce résultat met encore une fois en évidence la qualité de codage de la parole inférieure dans le réseau téléphonique cellulaire que dans le réseau téléphonique commuté, et confirme qu'il est nécessaire de réaliser tous les enregistrements dans un réseau homogène, commuté ou cellulaire. Cette contrainte a pour conséquence de devoir utiliser une base de données enregistrée dans le même type de réseau pour l'évaluation de la variabilité interlocuteur, ce qui peut se révéler un exercice difficile, car les bases de données existantes enregistrées par l'intermédiaire du réseau cellulaire sont encore peu nombreuses.

8.9. Évaluation de l'influence du bruit de fond

8.9.1. Procédure

La présence de bruit de fond dans les enregistrements de test est susceptible d'altérer les performances de la méthode de reconnaissance de locuteurs. L'influence de ce paramètre est quantifiée à l'aide d'enregistrements de test bruités artificiellement « Test 1-0dB », « Test 1-6dB », « Test 1-12dB », « Test 1-18dB », « Test 1-24dB » et « Test 1-30dB » et de l'enregistrement de test « Test 1 » non bruité, utilisé comme référence. Ces enregistrements ont été produits par addition à

l'enregistrement « Test 1 » d'un bruit de fond enregistré lors d'un apéritif dans une salle contenant une centaine de personnes.

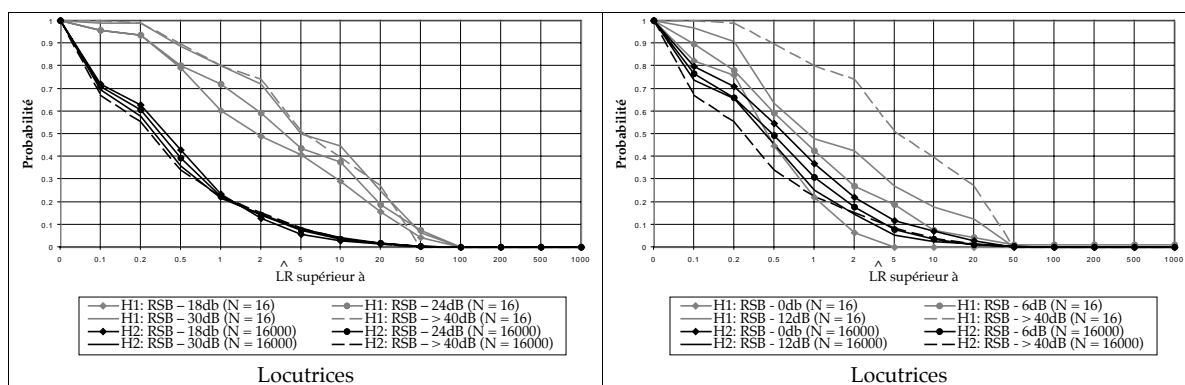
Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de test « Test 1-0dB », « Test 1-6dB », « Test 1-12dB », « Test 1-18dB », « Test 1-24dB » et « Test 1-30dB » aux modèles « Session Polyphone Cellulaire », « Session Comparaison » et « Polyphone 1 à 5 ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de test « Test 1-0dB », « Test 1-6dB », « Test 1-12dB », « Test 1-18dB », « Test 1-24dB » et « Test 1-30dB » avec les modèles de la voix des 1000 locutrices et des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.9.2. Résultats

Pour des raisons de lisibilité, les résultats des expériences effectuées avec les enregistrements de test « Test 1-0dB », « Test 1-6dB », « Test 1-12dB » et « Test 1 » sont présentés indépendamment des résultats des expériences effectuées avec les enregistrements de test « Test 1-18dB », « Test 1-24dB », « Test 1-30dB » et « Test 1 ».



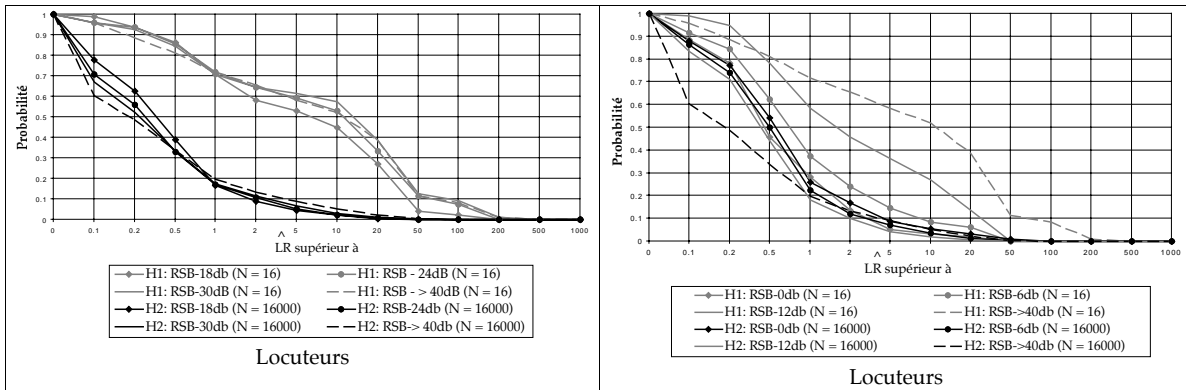


Figure VIII.15. Résultat de l'évaluation globale des rapports de vraisemblance en fonction du bruit de fond présent dans les enregistrements de test

8.9.3. Discussion des résultats

L'ajout de bruit de fond de manière artificielle ne permet pas de recréer des conditions parfaitement comparables à un enregistrement réalisé dans un environnement sonore bruité, notamment parce que dans un tel cas, le locuteur adapte son élocution aux caractéristiques de bruit de l'environnement ¹⁰⁵. Par contre cette manière de procéder permet de quantifier de manière précise le rapport signal sur bruit des enregistrements et de produire des tests comparables pour toutes les personnes. Le type de bruit ajouté à la parole a été choisi de manière à obtenir une situation réaliste ; il a été enregistré lors d'un apéritif dans une salle contenant une centaine de personnes, ce qui correspond au type de bruit que l'on peut retrouver dans un lieu public. Ce type de bruit additif est particulièrement défavorable, car il est lui-même constitué de parole.

Les résultats indiquent clairement que les performances sont inversement proportionnelles au niveau du bruit de fond. Un rapport signal-bruit de 18 dB semble la limite inférieure pour l'obtention d'un résultat exploitable avec la méthode utilisée et, à partir d'un rapport signal sur bruit de 6 dB, les rapports de vraisemblance mis en évidence lorsque l'hypothèse H_1 est vérifiée se confondent avec les rapports de vraisemblance mis en évidence lorsque l'hypothèse H_2 est vérifiée (Figure VIII.15.). L'avenir des méthodes de reconnaissance de locuteurs, en sciences forensiques notamment, passe par le développement de techniques de compensation efficaces permettant d'exploiter des signaux bruités, mais aucune technique universelle de compensation du bruit de fond n'a encore été proposée.

¹⁰⁵ *supra* : 2.3.4. Influence de la prise de son

8.10. Évaluation de l'influence du système d'enregistrement des indices

8.10.1. Procédure

L'influence du système utilisé pour l'enregistrement des indices est susceptible d'influencer les performances du système de reconnaissance automatique de locuteurs, surtout lorsque le système utilisé est un enregistreur analogique sur bande magnétique à faible vitesse de défilement. L'influence de ce paramètre est évaluée à l'aide des enregistrements de test « Test 1 » enregistré de manière numérique et « Test 1 analogique », enregistré sur l'enregistreur analogique de la Police Cantonale de Neuchâtel.

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de test « Test 1 » et « Test 1 analogique » au modèle « Session Comparaison ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat de la comparaison de ces mêmes enregistrements de test « Test 1 » et « Test 1 analogique » avec les modèles de la voix des 1000 locutrices et des 1000 locuteurs de la base de données « Polyphone Suisse Romande ».

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.10.2. Résultats

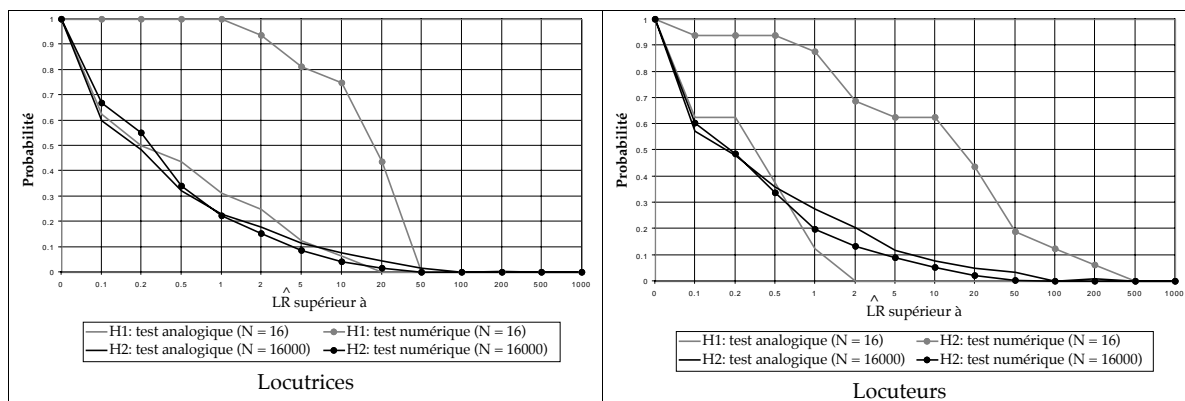


Figure VIII.16. Résultat de l'évaluation globale des rapports de vraisemblance en fonction du système utilisé pour l'enregistrement des indices

8.10.3. Discussion des résultats

Comme dans tous les domaines des sciences forensiques, les résultats montrent que la qualité de la procédure de récolte de l'indice conditionne entièrement les possibilités d'analyse subséquentes. Le signal de parole véhiculé par le réseau téléphonique contient une information ténue, mais utilisable pour la reconnaissance de locuteurs, les expériences précédentes le montrent. Or cette information est perdue par l'effet délétère des systèmes d'enregistrement analogiques et rien ne permet et ne permettra jamais de la régénérer. En effet l'utilisation d'un système d'enregistrement analogique altère tant le signal de parole que rien ne différencie plus les rapports de vraisemblance mis en évidence lorsque l'hypothèse H_1 est vérifiée des rapports de vraisemblance mis en évidence lorsque l'hypothèse H_2 est vérifiée (Figure VIII.16.).

Le passage à un format d'enregistrement numérique adéquat, sans compression du signal, est donc une condition *sine qua non* avant d'envisager une quelconque procédure d'expertise en reconnaissance de locuteurs à partir d'enregistrements recueillis par la police lors d'enquêtes. Cette situation représente malheureusement un handicap extrême pour l'implantation du système de reconnaissance de locuteurs mis au point, puisque toute démonstration de son efficacité dans une situation réelle est empêchée par la qualité intrinsèque des indices, enregistrés dans la grande majorité des cas avec des enregistreurs analogiques sur bande magnétique à faible vitesse de défilement.

8.11. Évaluation de l'influence de voix auditivement proches

Une hypothèse alternative parfois proposée par la personne suspectée est celle de l'existence d'une autre personne inconnue faisant partie de la population potentielle, dont la voix est si proche de la sienne qu'elles ne peuvent être différenciées par téléphone. Le cas de cette hypothèse alternative particulière a été testée dans plusieurs situations grâce aux 16 paires de personnes de la base de données « Polyphone IPSC », dont la voix est auditivement proche.

Cette configuration de test permet de comparer les rapports de vraisemblance qui peuvent être dégagés lorsque la personne mise en cause n'est pas la source de l'indice matériel, mais seulement une personne dont la voix est auditivement proche. Pour chaque paire de locutrices et de locuteurs, l'évaluation consiste à mettre en cause l'une des personnes de la paire en comparant ses propres enregistrements de test avec les modèles de sa voix d'une part et avec les modèles de voix de la seconde personne de la même paire d'autre part.

8.11.1. Influence du téléphone et de la ligne téléphonique

8.11.1.1. Procédure

Les personnes de la base de données « Polyphone IPSC » qui utilisent le même téléphone pour l'enregistrement des modèles et des enregistrements de test et celles qui utilisent des téléphones différents sont considérées de manière indépendante, afin d'évaluer l'influence du

paramètres du téléphone et de la ligne téléphonique lorsque les voix présentes dans le modèle et l'indice sont proches. Les tests ont été effectués à l'aide des sept modèles « Session Polyphone Cellulaire » « Session Comparaison » et « Session Polyphone 1 » à « Session Polyphone 5 » et enregistrements de test « Test 1 » à « Test 5 ».

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de test « Test 1 » à « Test 5 » avec ses propres modèles « Session Polyphone Cellulaire » « Session Comparaison » et « Session Polyphone 1 » à « Session Polyphone 5 ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de test « Test 1 » à « Test 5 » avec les sept modèles « Session Polyphone Cellulaire » « Session Comparaison » et « Session Polyphone 1 » à « Session Polyphone 5 » de la seconde personne de chaque paire de locutrices et de locuteurs.

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.11.1.2. Résultats

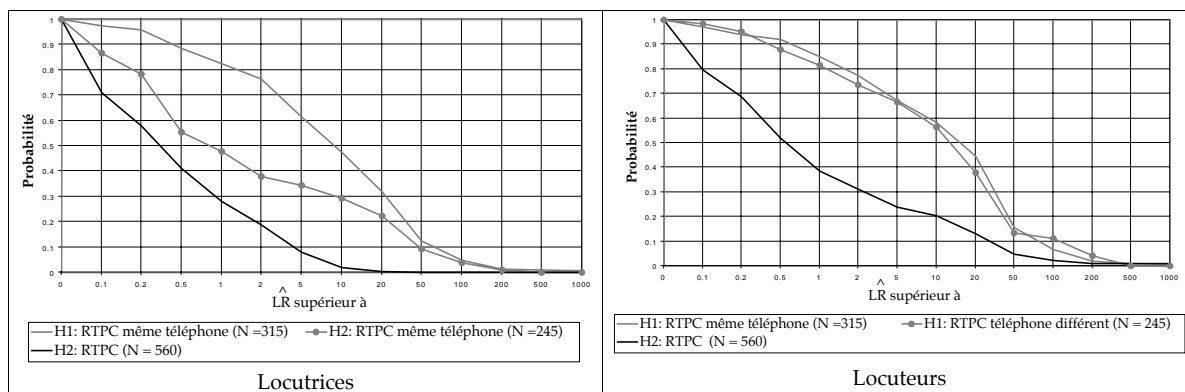


Figure VIII.17. Résultat de l'évaluation des rapports de vraisemblance en fonction du téléphone et de la ligne de téléphone utilisés, lorsque le locuteur est effectivement la source de l'enregistrement de test (H_1) et lorsqu'il s'agit d'une autre personne dont la voix est auditivement proche (H_2)

8.11.1.3. Discussion des résultats

Les résultats montrent encore une fois que la méthode de reconnaissance est plus robuste aux variations engendrées par le téléphone et la ligne téléphonique pour les locuteurs que pour les locutrices. Pour les locuteurs les résultats montrent qu'il est possible de distinguer la vraie source d'une personne dont la voix est auditivement proche. Pour les locutrices, cette différenciation est beaucoup plus difficile à faire si des téléphones différents ont été utilisés pour l'enregistrement des modèles et les enregistrements de test (Figure VIII.17).

Les résultats montrent aussi que la méthode est sensible au fait que les voix qui constituent le modèle et le test sont proches. En effet les rapports de vraisemblance mis en évidence sont plus élevés lorsque l'hypothèse H_2 équivaut à « une personne auditivement proche » (Figure VIII.17) que lorsque elle équivaut à « une personne de même langue, de même sexe et de même accent » (Figure VIII.10).

8.11.2. Influence du réseau téléphonique

8.11.2.1. Procédure

Le type de réseau téléphonique utilisé pour la production des enregistrements de test est susceptible d'influencer les performances du système automatique de reconnaissance de locuteurs lorsque les voix présentes dans le modèle et l'enregistrement de test sont proches. L'influence de ce paramètre a été évaluée à l'aide des enregistrements de test « Test cellulaire » et « Test 1 ».

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des tests « Test cellulaire » et « Test 1 » avec les sept modèles « Session Polyphone Cellulaire », « Session Comparaison » et « Session Polyphone 1 » à « Session Polyphone 5 ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des enregistrements de test « Test cellulaire » et « Test 1 » avec les sept modèles « Session Polyphone Cellulaire » « Session Comparaison » et « Session Polyphone 1 » à « Session Polyphone 5 » de la seconde personne de chaque paire de locutrices et de locuteurs.

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.11.2.2. Résultats

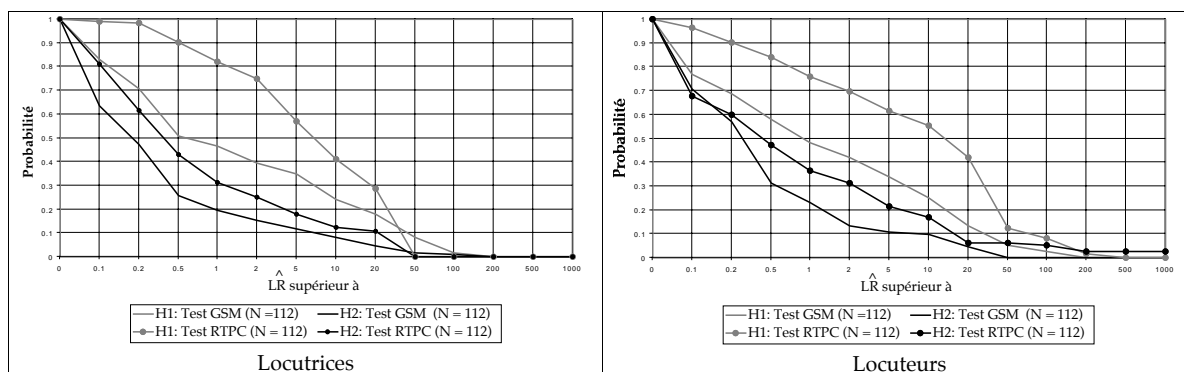


Figure VIII.18. Résultat de l'évaluation des rapports de vraisemblance en fonction du type de réseau téléphonique (GSM ou RTPC), lorsque le locuteur est effectivement la source de l'enregistrement de test (H_1) et lorsqu'il s'agit d'une autre personne dont la voix est auditivement proche (H_2)

8.11.2.3. Discussion des résultats

Les résultats montrent que l'utilisation du réseau cellulaire GSM pour la production des enregistrements de test altère les performances de la méthode par rapport à l'utilisation du réseau téléphonique commuté (RTPC) (Figure VIII.18).

D'un point de vue forensique, ce résultat montre qu'il est nécessaire de développer des méthodes d'extraction de caractéristiques dépendantes du locuteur spécialement adaptées aux algorithmes de codage utilisés dans le domaine de la téléphonie cellulaire. En effet, la plupart des indices soumis pour analyse sont aujourd'hui produits à partir de téléphones cellulaires, alors que les méthodes d'extraction de caractéristiques dépendantes du locuteur ont été développées pour les algorithmes de codage du réseau téléphonique commuté.

8.11.3. Influence d'un déguisement de la voix

8.11.3.1. Procédure

La présence d'un déguisement de la voix dans l'enregistrement de test est susceptible d'altérer les performances du système automatique de reconnaissance de locuteurs lorsque les voix présentes dans le modèle et l'indice sont proches. L'influence de ce paramètre est évaluée à l'aide des messages anonymes avec voix normale et déguisée « Test-an » et « Test-ad », enregistrés par chaque personne de la base de données « Polyphone IPSC ».

Dans la situation où l'hypothèse H_1 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des messages anonymes avec voix normale et déguisée « Test-an » et « Test-ad » avec les sept modèles « Session Polyphone Cellulaire », « Session Comparaison » et « Session Polyphone 1 » à « Session Polyphone 5 ».

Dans la situation où l'hypothèse H_2 est vérifiée, les éléments de preuve E sont le résultat, pour chaque personne de la base de données « Polyphone IPSC », de la comparaison des messages anonymes avec voix normale et déguisée « Test-an » et « Test-ad » avec les sept modèles « Session Polyphone Cellulaire », « Session Comparaison » et « Session Polyphone 1 » à « Session Polyphone 5 » de la seconde personne de chaque paire de locutrices et de locuteurs.

Les rapports de vraisemblance de ces éléments de preuve sont calculés de la manière suivante : le numérateur équivaut à la densité de probabilité de l'élément de preuve E dans la distribution de la variabilité intralocuteur du locuteur dont provient l'enregistrement analysé. Le dénominateur du rapport de vraisemblance équivaut à la densité de probabilité de l'élément de preuve E dans la distribution interlocuteur de l'enregistrement de test.

8.11.3.2. Résultats

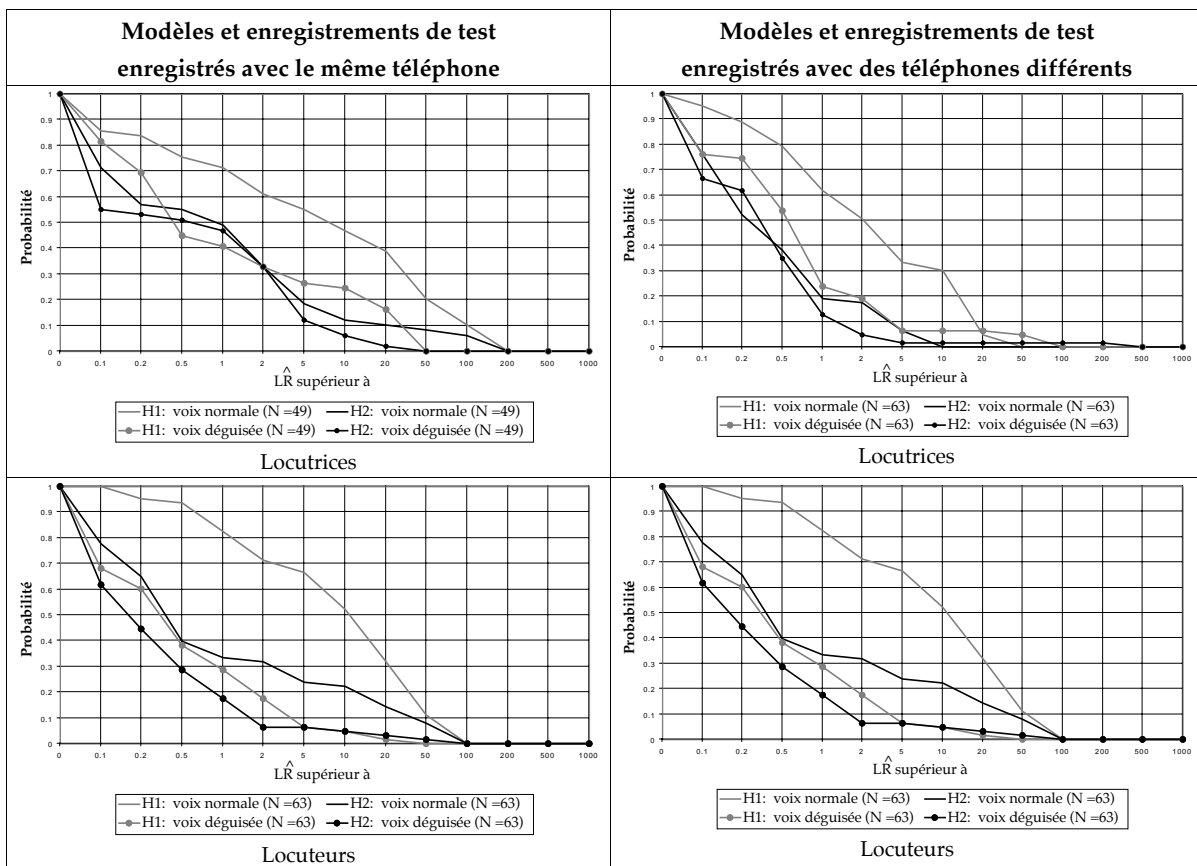


Figure VIII.19. Résultat de l'évaluation des rapports de vraisemblance en fonction de l'absence ou de la présence d'un déguisement dans l'enregistrement de test, lorsque le locuteur est effectivement la source de l'enregistrement de test (H_1) et lorsqu'il s'agit d'une autre personne dont la voix est auditivement proche (H_2)

8.11.3.3. Discussion des résultats

Les résultats impliquant les messages anonymes avec voix normale « Test-an » montrent des performances supérieures pour les locuteurs que pour les locutrices lorsque l'hypothèse H_1 est

vérifiée, ce qui confirme encore que le système est plus robuste pour les voix d'hommes que pour les voix de femmes (Figure VIII.19). Lorsque l'hypothèse H_2 est vérifiée, les rapports de vraisemblance mis en évidence sont nettement plus élevés si celle-ci équivaut à « une personne auditivement proche » (Figure VIII.19) que lorsqu'elle équivaut à « une personne de la même langue, de même sexe et de même accent », tant pour les locutrices que pour les locuteurs (Figure VIII.8).

Lorsque la voix des enregistrements de test est déguisée et qu'au départ il s'agit d'une voix auditivement proche, les rapports de vraisemblance mis en évidence lorsque l'hypothèse H_1 est vérifiée se confondent avec les rapports de vraisemblance mis en évidence lorsque l'hypothèse H_2 est vérifiée ». D'un point de vue forensique, ce résultat confirme que la méthode n'est pas utilisable en présence d'un déguisement de la voix dans l'indice.

8.11.4. Discussion sur les voix auditivement proches

L'hypothèse alternative que la personne mise en cause n'est pas la source de l'indice, mais seulement une personne dont la voix est auditivement proche, est la plupart du temps soulevée de manière judicieuse par la défense. Même si sa pertinence doit être examinée de cas en cas, la soulever suffit souvent à éveiller le doute. Lorsqu'une telle hypothèse alternative existe, l'utilisation du système de reconnaissance automatique de locuteurs nécessite une connaissance du type de réseau téléphonique par lequel a transité l'indice et du type de téléphone utilisé par l'auteur. Cette information peut être facilement recueillie, car le numéro de téléphone est transmis par le réseau téléphonique commuté numérique, tout comme par le réseau téléphonique cellulaire. Même si la corrélation entre numéro de téléphone et appareil téléphonique n'est pas parfaite, puisque plusieurs appareils peuvent être reliés à la même ligne sur le réseau commuté et que la carte *Subscriber Identity Module* (SIM) d'un téléphone portable est amovible, l'information reste utilisable, comme le pense aussi REYNOLDS [REYNOLDS, 1996].

PARTIE 4

SYNTHESE

IX. DISCUSSION GENERALE

9.1. Introduction

Le « principe d'individualité » de la voix n'est toujours qu'une hypothèse, preuve en est l'absence de caractéristique spécifique au locuteur mise en évidence à ce jour. Plusieurs indices laissent à penser que de nombreuses caractéristiques hautement discriminantes inhérentes à la voix restent à découvrir [NOLAN, 1990], mais la difficulté à décrire symboliquement et à définir l'information dépendante du locuteur dans la voix empêche une procédure réellement analytique, basée sur des descripteurs correspondant à des classes.

Cette réalité s'observe dans l'assise théorique lacunaire et controversée de la phonétique forensique et dans la difficulté à établir une méthodologie unifiée et satisfaisante pour la pratique de cette discipline. Ce même manque de connaissance théorique rend très partielle l'adéquation entre les stratégies de reconnaissance automatique et le modèle qu'elles sont censées décrire. La conséquence de ces insuffisances s'observe dans une robustesse toute relative des systèmes automatiques aux dégradations du signal de parole, qui peuvent être particulièrement importantes dans le domaine forensique.

9.2. Bilan de la recherche

9.2.1. Réflexion sur la démarche

L'estimation de rapports de vraisemblance, comme méthode d'inférence de l'identité du locuteur en sciences forensiques, remporte l'adhésion. Cependant, une clarification du rôle de l'expert et de sa démarche pour remplir son rôle de manière satisfaisante nous semble nécessaire, bien au-delà des solutions apportées par les différentes méthodes de reconnaissance, automatique et phonétique.

Les incertitudes rencontrées ne sont par essence pas différentes de celles rencontrées dans d'autres disciplines traitant quotidiennement avec l'incertitude, comme la médecine. Or ce domaine procède aussi par évaluation de rapports de vraisemblance. En effet, la pratique reconnue du diagnostic différentiel revient à mettre en concurrence la probabilité d'observer les symptômes (X) dans l'hypothèse de la présence d'une pathologie (Y) et la probabilité d'observer les mêmes symptômes dans l'hypothèse d'une pathologie alternative potentielle. Les hypothèses les plus pertinentes sont dégagées lors de la consultation du patient, sur la base d'observations et d'exams, et sont ensuite testées en laboratoire par des analyses adéquates. Dans l'environnement hospitalier, les résultats sont présentés sous forme de rapports de vraisemblance des hypothèses concurrentes, discutés lors d'un colloque quotidien qui rassemble l'ensemble des médecins spécialistes. Suite à la discussion, ceux-ci décident d'un traitement, jamais définitif, mais établi sur la base du bilan des connaissances à ce moment-là.

Dans le domaine forensique, l'évaluation de rapports de vraisemblance est déjà pratiquée dans les autres domaines de l'identification biométrique que sont l'analyse génétique [EVETT ET BUCKLETON, 1996] et la dactyloscopie [CHAMPOD, 1996], mais aussi dans certains domaines des microtraces, comme le verre [CURRAN *ET AL.*, 1998] et les fibres [ROUX, 1997]. Par analogie, le travail du criminaliste devrait aussi consister, dans le domaine de la reconnaissance de locuteurs, à mettre en concurrence la probabilité d'observer les caractéristiques de l'indice matériel X, dans l'hypothèse où il provient d'une source Y, et la probabilité d'observer les mêmes caractéristiques, dans l'hypothèse où l'indice matériel provient d'une source alternative potentielle. Les hypothèses les plus pertinentes sont, ou devraient être, dégagées au cours de l'enquête et sont, ou devraient être, testées ensuite en laboratoire par des analyses adéquates. Dans l'environnement du procès pénal, les résultats devraient donc aussi être présentés sous forme de rapports de vraisemblance des hypothèses concurrentes, de manière que la cour puisse les discuter lors des délibérations et prendre une décision de culpabilité ou d'innocence sur la base du bilan des connaissances à ce moment-là. L'adoption de cette méthode d'inférence de l'identité permet ensuite d'analyser l'aptitude des différentes méthodes proposées pour la reconnaissance de locuteurs en sciences forensiques à fournir des rapports de vraisemblance.

D'une part, cette approche n'est pas nouvelle en sciences forensiques, puisqu'elle avait déjà été considérée comme l'approche de choix par Poincaré dans l'affaire Francis Dreyfus [TARONI *ET AL.*, 1998]. D'autre part, son utilisation en sciences forensiques a été reconnue comme conforme à de nombreuses reprises, tant d'un point de vue logique que légal [LEWIS, 1984 ; EVETT, 1990 ; ROBERTSON ET VIGNAUX, 1995]. Pour reprendre l'observation de CHAMPOD dans le domaine de la dactyloscopie, les personnes en charge d'une expertise en reconnaissance de locuteurs « ne devraient jamais oublier qu'elles apportent à la cour un élément de preuve dont les principes d'interprétation sont équivalents à ceux de tout autre indice matériel » exploité en sciences forensiques [CHAMPOD, 1996].

Comme tout processus inductif, l'approche bayésienne ne permet pas d'aboutir à une conclusion ou à une décision sans recours à un principe par nature extrinsèque, sur la base de l'analyse de prémisses insuffisantes. C'est pourtant dans ces prémisses insuffisantes que réside la genèse même de la question qui nous occupe.

A notre avis, cette approche est apte à évaluer l'interprétation des résultats, tant dans une approche subjective, de type phonétique, que plus objective, de type automatique, avec l'avantage de mettre en lumière les apports et les limites de chacune des méthodes.

Finalement, la question de la validité scientifique d'une méthode, cataloguée comme telle, est une compétence de la communauté scientifique et non de l'autorité juridique, comme il a par exemple été prétendu dans *United States v Baller*¹⁰⁶, au sujet de la méthode spectrographique [GIANELLI ET IMWINKELRIED, 1986].

¹⁰⁶ [United States v Baller, (1975)519 F.2d 463,4th Cir. ; cert denied, 423 U.S. 1019]

9.2.2. Réflexion sur les méthodes

De notre point de vue, la portée du choix de la méthode de reconnaissance est moindre par rapport à celle de la démarche. La démarche choisie est en effet ouverte à toute approche, objective ou subjective, pour autant que sa validité dans l'évaluation de rapports de vraisemblance ait été démontrée.

9.2.2.1 Apports et limites de l'approche automatique

9.2.2.1.1. Choix de la méthode automatique

Pour être sélectionnée en vue d'une application forensique, une méthode de reconnaissance automatique de locuteurs doit satisfaire aux grands principes de l'indépendance par rapport au texte, fournir des résultats sous forme d'un nombre réel dans un ensemble de données continues, et être reconnue pour ses performances dans des conditions de dégradation du signal de la parole téléphonique.

9.2.2.1.2. Validité et fiabilité de la méthode automatique

Selon le principe de falsifiabilité, les expériences menées au cours de cette recherche ont permis de déterminer les principales conditions dans lesquelles la méthode automatique ne peut pas être appliquée, de manière à mieux cerner les conditions dans lesquelles elle est utilisable, sans pouvoir le démontrer toutefois.

La méthode automatique a l'avantage de permettre une évaluation de l'enregistrement considéré comme indice dans l'hypothèse où il provient de la personne mise en cause, sous la forme de la probabilité $p(E|H_1)$, utilisée comme numérateur dans le rapport de vraisemblance. Elle permet aussi une évaluation de l'enregistrement considéré comme indice, dans l'hypothèse où il ne provient pas de la personne mise en cause, sous la forme de la probabilité $p(E|H_2)$, utilisée comme dénominateur dans le rapport de vraisemblance; $p(E|H_2)$ indique la fréquence empirique des caractéristiques étudiées, dans la limite où l'échantillon dans lequel elles sont observées représente la population potentielle [LEWIS, 1984].

La puissance de calcul des micro-ordinateurs actuels permet d'envisager le problème sous un angle réellement statistique, en exploitant des bases de données de taille convenable pour la modélisation d'une population potentielle, le tout dans un temps raisonnable. En conséquence, l'évaluation empirique d'une méthode de reconnaissance automatique statistique dans des conditions forensiques est réalisable et permet de procéder à l'évaluation et à la quantification de ses performances. En cela, elle remplit le critère de recevabilité énoncé dans la *Federal Rule of Evidence* 702. De plus, l'évaluation de rapports de vraisemblance permet une bonne appréciation du coût de l'erreur, en fonction de l'évolution de la probabilité *a priori*.

Paradoxalement, le résultat de la quantification des performances représente aussi un frein psychologique à l'utilisation, en sciences forensiques, de toute méthode ne présentant pas des taux d'erreur théoriques infinitésimaux. Dans le domaine de l'analyse génétique, par exemple, l'argument d'un taux d'erreur infinitésimal est contesté et contestable lorsque la méthode est mise en pratique. LEMPERT estime d'ailleurs que le taux d'égale erreur résultant de l'analyse génétique se situe aux alentours de 2%, après la prise en compte de toutes les sources d'erreur, notamment celles de laboratoire [LEMPERT, 1995].

L'avancée constante et à un rythme soutenu de la recherche offre une marge de progression confortable aux méthodes de reconnaissance automatiques de locuteurs. Les capacités des classificateurs progressent en exploitant les solutions les plus évoluées dans les domaines de la reconnaissance de formes et de l'intelligence artificielle perceptive. A terme, cette marge de progression est cependant limitée en sciences forensiques si de nouvelles stratégies d'extraction des caractéristiques dépendantes du locuteur ne sont pas développées, par exemple sur la base d'un traitement local des distorsions présentes dans le signal de parole ou d'une ségrégation des parties intéressantes, par segmentation du signal de parole lors du prétraitement.

Pour l'instant, les critères de sélection des caractéristiques, définis par exemple par KWAN, ne sont que très partiellement satisfaits [KWAN, 1977 ; ROSENBERG ET SOONG, 1991 ; FURUI, 1994]. Cette constatation illustre parfaitement l'observation de BREMERMAN :

« un choix judicieux des caractéristiques conditionne plus de la moitié de l'efficacité de l'identification et aucun traitement mathématique postérieur ne saurait combler des caractéristiques mal choisies » [BREMERMAN, 1971 IN : KWAN, 1977].

Les expériences menées au cours de cette recherche à l'aide de plusieurs types d'enregistrements, utilisés pour la modélisation, le calcul de l'intravariabilité et de l'intervariabilité ainsi que pour les tests, donnent une première image des conditions nécessaires à l'obtention d'une réponse fiable de la part du système automatique développé. Les résultats montrent clairement que l'accession à un degré de fiabilité acceptable implique pour l'instant certaines restrictions d'utilisation, l'aménagement de dispositions techniques préalables et l'acceptation de la mise en évidence de rapports de vraisemblance modestes, manifestation des limites de la technologie actuelle.

Finalement, l'approche automatique et statistique de la reconnaissance de locuteurs est limitée à la prise en compte de l'information du signal de parole et cantonnée à l'opérationnalisation et à l'analyse de descripteurs objectivement mesurables. Elle ne peut espérer sauvegarder la richesse d'une analyse subjective, qui prend aussi en compte l'information de types linguistique, phonétique et dialectologique.

9.2.2.1.3. Compétences nécessaires à l'utilisation d'une méthode automatique

Comme le relève pertinemment KÜNZEL, le plus sophistiqué des systèmes nécessite l'interaction d'un expert à de nombreuses reprises, en commençant par la sélection d'énoncés de paroles adéquats [KÜNZEL, 1994A]. La méthode développée dans cette recherche tend cependant à limiter le travail et le facteur humain, en ne sollicitant le criminaliste que pour des activités simples,

qui ne nécessitent ni don ni habilité particulière, comme la segmentation des énoncés en phrases. Le choix de la méthode, comme celui de la base de données nécessaire au calcul de la variabilité interlocuteur et la composition des enregistrements de comparaison restent forcément en partie subjectifs, mais l'évaluation empirique est là pour guider ces choix et les justifier.

9.2.2.1.4. Indépendance par rapport aux langues analysées

Bien que l'indépendance des méthodes automatiques par rapport à la langue présente dans l'enregistrement d'indice ne soit pour le moment pas établie, l'application de ces méthodes n'est théoriquement pas limitée à une langue ou à un groupe de langues, si des bases de données nécessaires à la modélisation des différentes populations potentielles peuvent être trouvées sur le marché ou collectées. Notre courte expérience montre qu'en Suisse, les enregistrements présentés lors de demandes d'expertise en reconnaissance de locuteurs proviennent presque exclusivement d'écoutes téléphoniques réalisées au cours d'enquêtes sur le trafic de drogues illicites. Les langues parlées dans ces enregistrements dépendent fortement des ethnies qui noyautent le marché des drogues illicites, langues pour lesquelles il n'existe en général aucune base de données commerciale ou de référence.

9.2.2.2. Apports et limites de l'approche phonétique

9.2.2.2.1. Choix de la méthode auditive perceptive et phonétique acoustique

La voix humaine est tout d'abord un comportement et l'approche subjective rend possible la prise en compte d'une multitude de détails imparfaitement reconnus et difficiles à définir ou à cataloguer. Cette constatation est d'autant plus vraie que les descripteurs de l'identité véhiculés par la voix humaine sont encore largement inconnus.

9.2.2.2.2. Validité et fiabilité

L'approche phonétique peut être considérée comme « asymétrique ». En effet, cette méthode essentiellement comparative repose sur une grille d'analyse phonétique, dialectologique et linguistique. Elle permet d'attribuer de manière valide une probabilité subjective que les caractéristiques de l'indice matériel enregistré proviennent de l'enregistrement de comparaison, dans l'hypothèse où la personne mise en cause est le véritable auteur de l'enregistrement présenté comme indice, $p(E|H_1)$.

Lorsque le résultat de l'analyse indique de fortes dissemblances et de faibles ressemblances, donc lorsque le numérateur du rapport de vraisemblance est faible ou très faible, une évaluation grossière de la probabilité subjective que les caractéristiques de l'indice matériel enregistré proviennent d'une autre personne de la population potentielle, $p(E|H_2)$, peut être considérée comme acceptable. Dans ce cas, elle suffit à dégager un rapport de vraisemblance largement inférieur à 1.

Par contre, lorsque le résultat de l'analyse indique de faibles dissemblances et de fortes ressemblances, donc que le numérateur du rapport de vraisemblance est proche de 1, l'expérience de l'expert phonéticien ne peut pas être considérée comme un moyen acceptable pour inférer un dénominateur largement inférieur à 1, car seule une démarche statistique permet d'établir

empiriquement que la fréquence relative d'apparition des caractéristiques étudiées dans la population potentielle est extrêmement faible. Or, à l'exception des informations sur la fréquence fondamentale, le rythme et la puissance de la voix, ces statistiques de distribution sont inexistantes dans le domaine de la phonétique forensique [KÜNZEL *ET AL.*, 1995]. L'établissement de la distribution des caractéristiques analysées nécessite une base de données de la population potentielle et représente, pour chaque population potentielle, un travail conséquent pour une base de données de taille statistiquement valide.

Cette analyse du mécanisme de l'inférence de l'identité illustre les raisons pour lesquelles la validité des inférences de non-identité, fournies par l'approche phonétique, peut être considérée comme acceptable, alors que la validité des inférences d'identité ne l'est pas. En effet, l'apparente validité de telles inférences d'identité repose plus sur la valeur souvent élevée de la probabilité *a priori* dans ce domaine, que sur l'aptitude de l'analyse à dégager des rapports de vraisemblance très supérieurs à 1.

De nombreuses procédures d'identification reposant sur une approche exclusivement comparative souffrent de cette même invalidité. C'est notamment le cas pour l'identification d'écritures manuscrites, l'identification d'armes à feu à l'origine d'un projectile et l'identification de traces d'outils par comparaison visuelle de stries et de microstries. Malgré cette similitude, le degré de validité de chacune de ces approches doit être considéré comme différent. En effet, en l'absence de base de données, l'évaluation du dénominateur du rapport de vraisemblance fait largement appel aux capacités mnémoniques et à l'expérience de l'expert, qui peuvent être très différentes selon le type de perception sollicitée et la personne réalisant le travail.

« Malgré l'amélioration constante de la technologie utilisée par les phonéticiens dans leurs analyses forensiques, les conclusions qu'ils atteignent demeurent du niveau de l'opinion, et devraient être utilisées de façon corroborative » [FRENCH, 1994].

Cette réserve affichée par FRENCH ne prend pas en compte le fait que l'expertise en reconnaissance de locuteurs est souvent considérée comme une « *ultima ratio* », lorsque toutes les autres voies d'investigation ont été épuisées ou lorsque la voix enregistrée représente le seul lien entre l'auteur et l'infraction. L'observation de FRENCH semble aussi ignorer les contraintes de l'instruction pénale, qui obligent le magistrat à rechercher des preuves rapides et faciles à obtenir, comme les témoignages ou les aveux [GALLUSSER, 1998]. Dès lors, une procédure d'expertise en reconnaissance de locuteurs est décidée par le magistrat dans le but, non pas d'obtenir une preuve corroborative, mais une preuve centrale pour son dossier de renvoi devant un tribunal.

L'impression que ce type de preuve est utilisée de manière centrale est corroborée par la polémique qui a entouré l'étude du FBI de 1986. Sur 2000 cas d'identification de la voix répartis sur une période d'une quinzaine d'année, KOENIG prétend qu'une seule fausse identification (0,05%) a été comptabilisée avec la méthode spectrographique [KOENIG, 1986A]. Or sa méthodologie a été sévèrement critiquée par SHIPP, qui relève principalement que la supposition qu'une décision d'identification est correcte lorsqu'elle est compatible avec l'issue du cas est fautive [SHIPP *ET AL.*, 1987]. Les auteurs affirment avec raison qu'un critère tel qu'une décision de culpabilité ou d'innocence n'est pas suffisant pour établir la rectitude des décisions d'identification. Cette étude

illustre plutôt l'impact déterminant des résultats de l'analyse en reconnaissance de locuteurs sur l'issue du cas, alors que la méthode spectrographique n'est officiellement utilisée qu'à des fins d'enquête.

De plus, lorsque le magistrat entame une procédure en reconnaissance de locuteurs, il instruit dans la très forte majorité des cas à charge et non à décharge, situation pour laquelle l'approche phonétique n'est pas encore prête à donner des réponses valides.

Finalement l'appréciation des capacités des experts, bien qu'absolument nécessaire, reste extrêmement difficile à mettre sur pied, pour plusieurs raisons : le temps nécessaire à l'analyse d'un seul cas, le faible nombre de personnes actives dans la même langue, la constitution de cas fictifs de difficulté comparable dans des langues différentes ainsi que l'absence de consensus et d'unité autour des procédures d'analyse.

9.2.2.2.3. Indépendance par rapport aux langues analysées

L'article 6 (a) du code de procédure de l'IAFP (Annexe V.) souligne « que les membres devraient approcher avec la plus grande prudence l'analyse forensique d'échantillons de parole énoncés dans une autre langue que leur langue maternelle ».

Comme déjà mentionné, les langues parlées dans ces enregistrements dépendent fortement des ethnies qui noyautent le marché des drogues illicites. Dès lors, la formation d'experts phonéticiens compétents dans ces langues est d'une part presque utopique, ou en tout cas extrêmement longue, et d'autre part tout changement de situation géopolitique amenant une restructuration des canaux clandestins de distribution des drogues illicites peut rapidement rendre obsolètes les connaissances et l'expérience acquises.

9.2.3. Situation de l'approche spectrographique

La validité de l'approche de comparaison visuelle de spectrogrammes, telle qu'elle est pratiquée aux États-Unis, est contestable et contestée, tant par le vide théorique qui la caractérise, que par la controverse qu'a inauguré son application dans le domaine forensique. Malgré de multiples prises de position de scientifiques renommés, un rapport de l'Académie Nationale des Sciences des États-Unis et un certain nombre d'arrêts de la Cour Suprême, dans la foulée de l'arrêt Daubert, le système juridique nord-américain est malheureusement toujours incapable de se déterminer de manière unanime et définitive, alors qu'en 1946 déjà, les inventeurs du spectrographe s'étaient prononcés contre la validité de la méthode spectrographique pour la reconnaissance de locuteurs:

« It is axiomatic that no two individuals have voices that are exactly alike in pitch or vocal quality, but visible speech, in the form considered in this paper (spectrogrammes à bande étroite), does not emphasize these variables » [KOPP ET GREEN, 1946].

9.2.4. Réflexion sur les résultats

9.2.4.1. Choix du mode de présentation

Les variables qui influencent les performances d'un système de reconnaissance de locuteurs sont connues. L'originalité des résultats présentés dans cette recherche ne réside donc pas dans la mise en évidence de ces variables, mais plutôt dans la quantification de leur influence et dans la présentation des résultats sous une forme qui met en évidence l'évolution des rapports de vraisemblance.

9.2.4.2. Résultats de la procédure d'évaluation

L'évaluation des conditions requises pour l'application d'une approche automatique dans des conditions forensiques a deux utilités. La première concerne la possibilité d'établir *a priori* les chances de succès d'une procédure d'expertise en reconnaissance de locuteurs avec la méthode développée par une analyse préliminaire des caractéristiques de l'indice. La seconde se rapporte à l'utilisation du programme d'évaluation établi, comme base de comparaison de performances de différentes méthodes de reconnaissance sélectionnées pour l'application forensique.

9.2.4.2.1. Analyse préliminaire de l'indice

La procédure d'évaluation a montré la gradation de l'influence des variables sur les performances du système de reconnaissance développé. Jusqu'à deux ou trois mois, le temps qui sépare l'enregistrement de l'indice de celui du modèle n'a qu'une influence moyenne sur la qualité des résultats ; le contenu de l'enregistrement utilisé comme modèle n'a pas une importance capitale non plus, ce qui montre le caractère indépendant du texte de la méthode.

La qualité et la quantité des données qui constituent l'indice ont par contre une grande influence. Quatre secondes de parole sont à considérer comme un minimum pour l'obtention d'un résultat acceptable, et la présence d'un déguisement quel qu'il soit détériore les résultats de manière prépondérante.

Des caractéristiques techniques très différentes des téléphones et des lignes téléphoniques utilisés pour l'enregistrement de l'indice et l'enregistrement de comparaison peuvent aussi avoir une influence négative importante sur les résultats : des conditions hétérogènes comme une différence de technologie des microphones équipant les téléphones ou la différence entre les algorithmes de codage utilisés pour le réseau téléphonique public commuté numérique et le réseau cellulaire numérique peuvent engendrer une sérieuse diminution de performance.

La présence de bruit de fond dans l'indice a un effet négatif sur les performances du système. Cet effet dépend du rapport entre l'intensité du bruit de fond et l'intensité du signal de parole, mais aussi de la nature de ce bruit de fond. Avec un bruit de fond constitué de parole humaine, tel qu'il existe dans les lieux publics, un rapport signal sur bruit de 18 dB semble la limite inférieure pour l'obtention d'un résultat exploitable.

Le type de matériel d'enregistrement utilisé pour la collecte de l'indice et de l'enregistrement de comparaison est prépondérant ; avec le système de reconnaissance développé, seule l'utilisation

d'un équipement d'enregistrement numérique, capable d'enregistrer directement l'information véhiculée par le réseau téléphonique, engendre des résultats utilisables.

Finalement, la composition de la population potentielle des auteurs de l'indice peut avoir une importance essentielle sur les performances. En effet, comme les caractéristiques analysées ne sont pas spécifiques, mais seulement dépendantes du locuteur, elles sont aussi influencées par le système de transmission. Les performances peuvent notamment s'amenuiser lorsque la population potentielle est réduite à une personne dont la voix est proche de celle de l'auteur et que celle-ci est enregistrée dans une configuration téléphonique comparable, par exemple avec le même téléphone, par la même ligne téléphonique ou par l'intermédiaire de la même cellule d'un réseau cellulaire.

9.2.4.2.2. Le programme d'évaluation comme base de comparaison de méthodes différentes

Le programme d'évaluation réalisé dans le cadre de cette recherche permet d'observer les possibilités et les limites du système de reconnaissance. Le classificateur par modélisation de mélanges de fonctions de densité gaussiennes peut être considéré comme un outil générique utilisable à moyen terme ; par contre l'évaluation d'algorithmes d'analyse du signal de parole particulièrement adaptés aux différentes conditions rencontrées en sciences forensiques peut être envisagée dès aujourd'hui, sur la base du programme d'évaluation réalisé.

9.2.5. Voies de recherche

9.2.5.1. Recherche fondamentale

Les principales questions non résolues dans le domaine de la reconnaissance automatique de locuteurs ont aussi été répertoriées par FURUI [FURUI, 1997]. Elles concernent premièrement l'apport de connaissances dans le domaine des processus de production de la parole et de la reconnaissance de locuteurs par les êtres humains ; deuxièmement elles concernent les caractéristiques dépendantes du locuteur et finalement la modélisation de la variabilité interlocuteur et intralocuteur, notamment la variabilité intralocuteur à long terme et le déguisement.

En 1983 déjà, NOLAN mentionnait un certain désintérêt pour la recherche fondamentale dans le domaine de la variabilité interlocuteur :

« Un manque d'intérêt pour la complexité de la base des différences interlocuteur, que ce soit par ignorance de cette complexité ou par confiance exagérée en une sophistication technologique et statistique toujours plus grande, laisse ceux qui recommandent une application pratique des projets de reconnaissance du locuteur ouverts à une sérieuse critique théorique » [NOLAN, 1983].

En 1995, NOLAN précise sa pensée et ouvre de nouvelles voies de recherche en indiquant que l'analyse de la prosodie est généralement considérée comme accessoire, ignorée la plupart du temps et traitée non comme un aspect du système phonologique, mais purement comme un aspect non structuré du signal de parole, en termes de paramètres tels que la fréquence fondamentale perçue. Certains concepts et représentations phonologiques sont ostensiblement absents de la

reconnaissance de locuteurs, alors qu'un modèle phonologique de l'intonation permettrait la mise en évidence de phénomènes potentiellement spécifiques au locuteur [NOLAN, 1995].

L'enjeu des questions de FURUI et des propositions de NOLAN dépasse largement la problématique abordée ici, laquelle montre l'importance de la recherche fondamentale dans la quête de la compréhension des mécanismes de production et de perception de la parole, qui restent encore en grande partie à découvrir.

La levée de ces incertitudes serait en tout cas d'une grande utilité pour fonder l'assise théorique de la phonétique forensique et améliorer les stratégies de reconnaissance automatique de locuteurs lorsque le signal de parole est dégradé.

9.2.5.2. Recherche appliquée

Dans tout processus d'identification, la principale lacune concerne les bases de données. D'une part, il n'existe que très peu de bases de données qui prennent en compte des éléments de la variabilité intralocuteur rencontrés en sciences forensiques tels que le déguisement ou la variabilité intralocuteur à long terme. D'autre part, la modélisation des populations potentielles demeure difficile car, s'il est possible de modéliser la population suisse romande et alémanique de manière acceptable avec des bases de données comme Polyphone, les bases de données dans des langues pertinentes, du point de vue de l'enquête criminelle, sont rares ou inexistantes. A notre avis, une base de données à caractère forensique devrait, non seulement contenir un nombre suffisant de locuteurs pour modéliser la population potentielle, mais comporter aussi plusieurs enregistrements de chaque locuteur, enregistrés avec plusieurs appareils téléphoniques, dont une fraction importante de téléphones cellulaires, et sur une période aussi longue que possible, de l'ordre d'une année. Ce type de base de données pourrait être utilisé pour la modélisation des variabilités intralocuteur et interlocuteur dans l'approche automatique ; il permettrait aussi l'établissement progressif de la distribution des caractéristiques analysées dans l'approche phonétique, contribuant ainsi à la valider. Alliant toutes ces qualités, les enregistrements réalisés lors des procédures d'écoute téléphonique seraient une source idéale pour constituer des bases de données dans les langues diverses et intéressantes dans l'investigation.

9.3. Utilisation dans la réalité de l'approche automatique développée

9.3.1. Aspects méthodologiques

Cette recherche a contribué à clarifier l'aspect méthodologique des approches subjectives et objectives pratiquées dans le domaine de la reconnaissance de locuteurs en sciences forensiques. L'exigence principale avant une utilisation consiste dans la possibilité de tester la méthode *in situ*, car malgré la volonté de calquer la réalité forensique au plus près, la procédure d'évaluation menée dans cette recherche doit être considérée comme une évaluation *in vitro*.

9.3.2. Aspects techniques

Les résultats présentés montrent que l'information dépendante du locuteur contenue dans le signal de parole téléphonique est perdue lors de son enregistrement sur un support analogique de mauvaise qualité. Le passage à l'enregistrement dans un format numérique adéquat, sans compression du signal, est une condition *sine qua non* avant d'envisager une quelconque procédure d'expertise en reconnaissance de locuteurs à partir d'enregistrements recueillis lors d'enquêtes de police. Les systèmes d'acquisition et d'édition numérique assistés par ordinateur font actuellement quasiment partie des applications grand public, le CD devient un support numérique quasi universel tant pour les données audio qu'informatiques et le prix de revient du support de type *Recordable Compact Disc* (CD-R) est largement inférieur à celui d'une cassette audio compact de qualité. De plus, la pérennité et l'intégrité de l'information sont garanties sur CD-R, puisque celui-ci ne peut plus être modifié une fois gravé. Ce passage est d'autant plus nécessaire que, dans certaines affaires importantes récentes, comme l'affaire du mafieux russe présumé Mikhaïlov, les enregistrements téléphoniques constituaient une des pièces maîtresses de l'accusation, mais leur qualité déplorable n'a permis aucune expertise en reconnaissance de locuteurs [GUELPA ET SCHAAD, 1998].

Une deuxième mesure technique essentielle consisterait à séparer les signaux provenant des différents interlocuteurs et l'enregistrement de chacune des voix sur une piste séparée en cas de dialogue ou de conversation entre plusieurs personnes. Cette mesure éviterait d'une part toute procédure de ségrégation des locuteurs, manuelle ou automatique, et faciliterait grandement le travail de retranscription des conversations téléphoniques en permettant une écoute indépendante des interlocuteurs. Finalement cette mesure empêcherait la contamination d'une piste par un bruit de fond éventuel présent sur l'autre.

9.3.3. Aspects juridiques

Les enregistrements effectués durant des procédures d'écoute téléphonique seraient une excellente source pour la constitution des bases de données pour des langues inexistantes sur le marché. Cependant cette solution passe par un contrôle du respect des lois et plus particulièrement de l'art. 13 al. 1 de la Constitution Fédérale (CF) du 18 décembre 1998 et de la loi sur la protection des données.

L'information essentielle à fournir, pour qu'une décision favorable puisse être obtenue, consiste à expliquer et à démontrer aux autorités de surveillance que, pour chaque locuteur, seuls les paramètres utiles pour la reconnaissance extraits des enregistrements de référence sont conservés et non l'enregistrement. Par analogie, dans le domaine de l'analyse génétique, seul le résultat concernant les *loci* analysés, et non tout le patrimoine génétique, est conservé dans la base de données de référence. De cette manière, les données sensibles ne feraient que transiter par la machine le temps de l'analyse, mais ne seraient pas stockées.

9.3.4. Aspects d'organisation

Notre expérience laisse supposer l'existence d'un « chiffre noir » important du nombre de cas nécessitant le recours à une expertise en reconnaissance de locuteurs en Suisse. L'idée qu'aucune possibilité technique de reconnaissance de locuteurs n'existe est véhiculée aux magistrats instructeurs par les policiers. Le quota des magistrats qui se satisfont d'une telle réponse et qui ne poursuivent pas leurs investigations ne peut pas être évalué. Par contre, depuis cinq ans, chaque année entre 10 et 20 magistrats et quelques avocats ont contacté l'Institut de Police Scientifique et de Criminologie de l'Université de Lausanne, demandant une solution dans le domaine de la reconnaissance de locuteurs, ce qui montre la nécessité et l'utilité d'une réponse dans ce domaine, car les affaires dont il est question concernent généralement des écoutes téléphoniques liées à d'importants trafics de drogues illicites.

Répondre à ce besoin nécessite le développement d'une infrastructure, qui pourrait être unique pour le pays au vu du volume d'affaires identifiées à l'heure actuelle. Sa place devrait sans aucun doute se trouver dans un laboratoire national de sciences forensiques, comme il en existe dans tous les pays d'Europe. Mais en Suisse, en l'absence d'une telle entité, il est actuellement difficile de déterminer la place idéale d'un laboratoire de reconnaissance de locuteurs et malheureusement, le fédéralisme suisse fragmente et cloisonne les services publics. Ce phénomène de *linkage blindness*, notamment décrit par RIBAUX dans le domaine de l'analyse criminelle, est confirmé par l'absence d'intérêt de la part des services fédéraux consultés dans le but d'établir des contacts sur la question de la reconnaissance de locuteurs en sciences forensiques ; ils sont pourtant régulièrement touchés par ce problème [RIBAUX, 1997].

X. CONCLUSION

« Une machine devrait-elle être construite pour montrer qu'à partir d'un ensemble de locuteurs, elle réalise l'identification de locuteurs mieux que les auditeurs humains ? »
[LEWIS, 1984].

La question est courte, mais la réponse ne l'est pas, comme le montre l'étude des différentes approches envisagées pour y répondre. Nous pensons avoir contribué à définir quantitativement les performances de l'être humain et de la machine, et pouvoir ainsi répondre à la question de LEWIS.

Sur un plan théorique, la démarche par évaluation de rapports de vraisemblance est conforme d'un point de vue logique, et elle permet l'interprétation des résultats obtenus tant par des méthodes subjectives qu'objectives. Néanmoins, seule l'approche automatique possède actuellement la capacité d'appréhender la question sous un angle réellement statistique.

Sur un plan pratique, l'approche automatique est toujours considérée comme expérimentale dans la plupart des pays. Les résultats de cette recherche montrent que les rapports de vraisemblance dégagés sont encore modestes, mais ils ont l'avantage de reposer sur une méthodologie valide. D'autre part, les valeurs de rapports de vraisemblance mises en évidence ne demandent qu'à évoluer en fonction des progrès technologiques dans les domaines de la collecte de l'indice matériel et de la reconnaissance automatique de locuteurs.

Si l'approche phonétique comparative peut être considérée comme valide dans une démarche d'inférence de la non-identité du locuteur, sa validité dans la démarche d'inférence de l'identité du locuteur sera contestable et contestée tant qu'aucune statistique fiable de la distribution des caractéristiques analysées dans la population potentielle n'aura pas été établie. Sur un plan pratique, l'approche phonétique est acceptée dans certains pays, mais rejetée dans d'autres. La validité de la reconnaissance auditive de locuteurs par des profanes n'a qu'une valeur comparable à celle d'un autre témoignage.

Finalement, la validité de la méthode spectrographique, basée sur la comparaison visuelle de spectrogrammes vocaux, est contestable et contestée, tant par le vide théorique qui la caractérise, que par la controverse qu'a soulevée son application dans le domaine forensique. Sur un plan pratique, elle est de moins en moins pratiquée, mais subsiste encore dans certains États des États-Unis.

« Et de même que l'écriture n'est pas la même chez tous les hommes, les mots parlés ne sont pas non plus les mêmes » [ARISTOTE, 384 - 322 av. J.-C.].

Près de 2400 ans après la naissance d'ARISTOTE, personne n'a encore, à notre connaissance, relevé le défi de la démonstration de l'individualité de la voix humaine par rapport à la population de la Terre, ni celui de l'écriture manuscrite d'ailleurs. À l'aube du troisième millénaire, l'individualité de la voix humaine demeure donc une hypothèse.

ANNEXES

ANNEXE I. EXTRAITS DE LA CONSTITUTION FEDERALE DE LA CONFEDERATION SUISSE (RS 101)

du 29 mai 1874 (État le 20 avril 1999)

Chapitre premier: Dispositions générales

Art. 36

¹Dans toute la Suisse, les postes et les télégraphes sont du domaine fédéral.

²Le produit des postes et des télégraphes appartient à la caisse fédérale.

³Les tarifs seront fixés d'après les mêmes principes et aussi équitablement que possible dans toutes les parties de la Suisse.

⁴L'inviolabilité du secret des lettres et des télégrammes est garantie.

Mise à jour du 18 décembre 1998, adoptée par le peuple suisse le 6 juin 1999

(État le 26 octobre 1999)

Titre 2: Droits fondamentaux, citoyenneté et buts sociaux

Chapitre premier: Droits fondamentaux

Art. 13 Protection de la sphère privée

¹Toute personne a droit au respect de sa vie privée et familiale, de son domicile, de sa correspondance et des relations qu'elle établit par la poste et les télécommunications.

²Toute personne a le droit d'être protégée contre l'emploi abusif des données qui la concernent.

ANNEXE II. EXTRAITS DU CODE PENAL SUISSE (RS 311.0)

du 21 décembre 1937 (État le 10 novembre 1998)

Livre premier: Dispositions générales

Première partie: Des crimes et des délits

Titre deuxième: Conditions de la répression

Art. 33 (Légitime défense)

¹Celui qui est attaqué sans droit ou menacé sans droit d'une attaque imminente a le droit de repousser l'attaque par des moyens proportionnés aux circonstances ; le même droit appartient aux tiers.

²Si celui qui repousse une attaque a excédé les bornes de la légitime défense, le juge atténuera librement la peine (art. 66) ; si cet excès provient d'un état excusable d'excitation ou de saisissement causé par l'attaque, aucune peine ne sera encourue.

Art. 34 (État de nécessité)

¹Lorsqu'un acte aura été commis pour préserver d'un danger imminent et impossible à détourner autrement un bien appartenant à l'auteur de l'acte, notamment la vie, l'intégrité corporelle, la liberté, l'honneur, le patrimoine, cet acte ne sera pas punissable si le danger n'était pas imputable à une faute de son auteur et si, dans les circonstances où l'acte a été commis, le sacrifice du bien menacé ne pouvait être raisonnablement exigé de l'auteur de l'acte.

Si le danger était imputable à une faute de ce dernier ou si, dans les circonstances où l'acte a été commis, le sacrifice du bien menacé pouvait être raisonnablement exigé de l'auteur de l'acte, le juge atténuera librement la peine (art. 66).

²Lorsqu'un acte aura été commis pour préserver d'un danger imminent et impossible à détourner autrement un bien appartenant à autrui, notamment la vie, l'intégrité corporelle, la liberté, l'honneur, le patrimoine, cet acte ne sera pas punissable. Si l'auteur pouvait se rendre compte que le sacrifice du bien menacé pouvait être raisonnablement exigé de celui auquel le bien appartenait, le juge atténuera librement la peine (art. 66).

Livre deuxième: Dispositions spéciales

Titre troisième:

Infractions contre l'honneur et contre le domaine secret ou le domaine privé

Art. 179 ter (Enregistrement non autorisé de conversations)

Celui qui, sans le consentement des autres interlocuteurs, aura enregistré sur un porteur de son une conversation non publique à laquelle il prenait part, celui qui aura conservé un enregistrement qu'il savait ou devait présumer avoir été réalisé au moyen d'une infraction visée au premier alinéa, ou en aura tiré profit, ou l'aura rendu accessible à un tiers, sera, sur plainte, puni de l'emprisonnement pour un an au plus ou de l'amende.

Art. 179 septies (Abus de téléphone)

Celui qui, par méchanceté ou par espièglerie, aura abusé d'une installation téléphonique soumise à la régale des téléphones pour inquiéter un tiers ou pour l'importuner sera, sur plainte, puni des arrêts ou de l'amende.

Art. 179 octies (Mesures officielles de surveillance)

¹N'est pas punissable celui qui, dans l'exercice d'une attribution que lui confère expressément la loi, ordonne des mesures officielles de surveillance de la correspondance postale et des télécommunications de personnes déterminées ou prescrit l'utilisation d'appareils techniques de surveillance (art. 179 bis et s.), à condition qu'il demande immédiatement l'approbation du juge compétent.

²L'approbation visée au 1^{er} alinéa peut être donnée aux fins de poursuivre ou de prévenir un crime ou un délit dont la gravité ou la particularité justifie l'intervention.

ANNEXE III. ORDONNANCE SUR LE SERVICE DE SURVEILLANCE DE LA CORRESPONDANCE POSTALE ET DES TELECOMMUNICATIONS (RS 780.11)

du 1^{er} décembre 1997 (État le 31 décembre 1997)

Le Conseil fédéral suisse,

vu l'article 43, 2^e alinéa, de la loi sur l'organisation du gouvernement et de l'administration ;

vu les articles 44 et 62 de la loi du 30 avril 1997 sur les télécommunications ;

vu l'article 4 de la loi fédérale du 4 octobre 1974 instituant des mesures destinées à améliorer les finances fédérales,

arrête:

Section 1: Organisation

Article premier (Principe)

La Confédération exploite un service chargé de surveiller la correspondance postale et les télécommunications (service).

Art. 2 (Subordination)

¹Le service est rattaché administrativement au Département fédéral de l'environnement, des transports, de l'énergie et de la communication (département).

²Il exécute ses tâches de manière autonome, sous la surveillance du département.

Art. 3 (Collaboration avec les autorités concédantes)

Le service accomplit ses tâches en collaboration avec les autorités concédantes et de surveillance actives dans le domaine des postes et des télécommunications.

Section 2: Surveillance de la correspondance postale

Art. 4 (Tâches du service)

¹En matière de correspondance postale, le service remplit les tâches suivantes:

- a. il s'assure que la surveillance soit conforme au droit applicable et qu'elle ait été ordonnée par une autorité compétente ;
- b. il ordonne à la Poste d'exécuter la surveillance ;
- c. il communique immédiatement la levée de la surveillance à l'autorité qui l'a approuvée ;
- d. il conserve l'ordre de surveillance durant une année après la levée de celle-ci.

²A la demande de l'autorité qui a ordonné la surveillance, le service peut lui fournir des conseils techniques en la matière.

³Le service demande à la Poste les informations nécessaires à la mise en œuvre de la surveillance.

Art. 5 (Obligations de la Poste)

¹La Poste exécute la surveillance conformément à l'ordre qu'elle a reçu et en contact direct avec l'autorité qui l'a ordonnée. Elle met à disposition les équipements nécessaires.

²Elle communique la levée de la surveillance au service.

³La surveillance et toutes les informations qui s'y rapportent sont soumises au secret postal et au secret des télécommunications (art. 33^{ter} CP).

Section 3: Surveillance des télécommunications

Art. 6 (Tâches du service)

¹En matière de surveillance des télécommunications, le service remplit les tâches suivantes:

- a. il s'assure que la surveillance soit conforme au droit applicable et qu'elle ait été ordonnée par une autorité compétente ;
- b. il ordonne aux fournisseurs de services de télécommunication de prendre les mesures nécessaires à l'exécution de la surveillance ;
- c. il reçoit les communications de la personne surveillée, déviées par les fournisseurs de services ; il les enregistre et les transmet à l'autorité qui a ordonné la surveillance ;
- d. il veille à l'installation de raccordements directs, mais il n'enregistre pas les communications qui ont lieu via ces derniers ;
- e. il reçoit les relevés de service des fournisseurs de services de télécommunication et les transmet à l'autorité qui a ordonné la surveillance ;
- f. il communique immédiatement la levée de la surveillance à l'autorité qui l'a approuvée ;
- g. il conserve l'ordre de surveillance durant une année après la levée de celle-ci.

²Les autorités qui ordonnent et approuvent la surveillance peuvent charger le service de:

- a. trier les communications enregistrées ;
- b. mettre en place des mesures de protection lorsque sont surveillés des tiers, des cabines téléphoniques publiques ou des personnes qui, selon le droit procédural applicable, peuvent refuser de témoigner car elles sont tenues au secret professionnel.

³Dans la mesure de ses capacités en personnel et en moyens techniques, le service peut également être chargé des tâches suivantes:

- a. enregistrer les communications effectuées sur les raccordements directs ;
- b. transcrire ces enregistrements ;
- c. traduire les transcriptions rédigées en langues étrangères ;

d. fournir des conseils techniques aux autorités et aux fournisseurs de services de télécommunication.

⁴Le service demande aux fournisseurs de services les informations nécessaires à la mise en œuvre de la surveillance.

Art. 7 (Obligations des fournisseurs de services de télécommunication)

¹A la demande du service, les fournisseurs de services de télécommunication sont tenus de lui transmettre les communications de la personne surveillée et les relevés de service ainsi que les informations nécessaires à la mise en œuvre de la surveillance.

²Ils fournissent dans les meilleurs délais les relevés de service demandés et transmettent si possible en temps réel les communications de la personne surveillée. Ils suppriment les cryptages.

³Ils mettent à disposition les équipements nécessaires à l'exécution de la surveillance.

⁴La surveillance et toutes les informations qui s'y rapportent sont soumises au secret postal et au secret des télécommunications (art. 321^{ter} CP).

Section 4: Renseignements sur les raccordements

Art. 8 (Tâches du service)

¹Pour les motifs suivants, le service fournit des renseignements sur les raccordements uniquement aux autorités suivantes, à leur demande:

- a. pour déterminer les raccordements et les personnes à surveiller: aux autorités fédérales et cantonales qui ordonnent ou approuvent la surveillance des télécommunications ;
- b. pour exécuter des tâches de police: à l'Office fédéral de la police, à la police fédérale, au service de sécurité de l'administration fédérale et aux commandements des polices cantonales et municipales ;
- c. pour régler des affaires relevant du droit pénal administratif: aux autorités fédérales et cantonales compétentes en la matière.

²Le service peut charger les fournisseurs de services de télécommunication de donner directement aux autorités les renseignements sur les raccordements de télécommunications.

³Le service conserve les demandes de renseignements pendant un an.

Art. 9 (Devoirs des fournisseurs de services de télécommunication)

¹A la demande du service, les fournisseurs de services de télécommunication lui fournissent les données suivantes sur les raccordements, pour autant qu'ils les possèdent:

- a. le nom, l'adresse et la profession de l'utilisateur ;
- b. les ressources d'adressage du raccordement selon l'article 3, lettre f, de la loi du 30 avril 1997 sur les télécommunications ;
- c. le type de raccordement.

²Le service peut également obtenir les informations prévues au 1^{er} alinéa en consultant directement des banques de données.

³Les fournisseurs de services de télécommunication mettent à disposition les équipements servant à obtenir ces renseignements.

Section 5: Dispositions communes

Art. 10 (Émoluments et indemnités)

¹Le département fixe:

- a. les émoluments pour les prestations du service prévues aux articles 4, 6 et 8 ;
- b. les indemnités pour les frais de la Poste et des fournisseurs de services de télécommunication.

²Le service adresse sa facture aux autorités qui ont ordonné la surveillance et établit le décompte des prestations de la Poste et des fournisseurs de services de télécommunication.

³Lorsque les renseignements sont fournis sans intermédiaire, les fournisseurs de services de télécommunication facturent les taxes directement aux autorités compétentes.

⁴Après un délai de trois ans au plus tard, les émoluments prévus au premier alinéa, lettre a, doivent couvrir les coûts.

Art. 11 Ordres de surveillance et demandes de renseignements

¹Les ordres de surveillance et les demandes de renseignements doivent être adressés par écrit ou par téléfax au service ou aux fournisseurs de services de télécommunication, qui sont mandatés pour donner directement les renseignements sur les raccordements de télécommunications.

²En cas d'urgence, les ordres de surveillance peuvent aussi être communiqués oralement. Toutefois, l'autorité ordonnant la surveillance ne recevra les communications de la personne surveillée, les relevés de service des fournisseurs de services de télécommunication ou les résultats de la surveillance de la correspondance postale d'une personne qu'après avoir confirmé son ordre par écrit ou par télécopie.

³Les ordres de surveillance doivent expressément mentionner les faits au sens du code pénal ou de tout autre acte législatif sur lesquels se fonde l'instruction ou l'infraction qu'il y a lieu de prévenir.

⁴Les compléments aux ordres de surveillance ainsi que les modifications et les prorogations de ces derniers doivent également être adressés au service par écrit ou par télécopie.

⁵En cas d'urgence, les demandes de renseignements concernant des raccordements aux services de télécommunication peuvent aussi être adressées oralement, mais elles doivent être immédiatement confirmées par écrit ou par téléfax. Les renseignements requis sont transmis par écrit ou par téléfax au service désigné par les autorités compétentes.

Section 6: Dispositions finales

Art. 12 **Exécution**

¹Le département applique la présente ordonnance.

²Il édicte les dispositions d'exécution relatives à l'organisation et aux tâches du service, aux obligations de la Poste et des fournisseurs de services de télécommunication ainsi qu'à la teneur minimale des ordres de surveillance.

Art. 13 **Entrée en vigueur et validité**

¹La présente ordonnance entre en vigueur le 1^{er} janvier 1998.

²Sa validité expire lors de l'entrée en vigueur d'une loi sur la surveillance de la poste et des télécommunications.

Annexe IV. Extraits des *Federal Rules of Evidence*

28 United States Code, Appendix current through 11/7/94

Article I : General Provisions

Rule 104. Preliminary Questions

(a) Questions of admissibility generally

Preliminary questions concerning the qualification of a person to be a witness, the existence of a privilege, or the admissibility of evidence shall be determined by the court, subject to the provisions of subdivision (b). In making its determination it is not bound by the rules of evidence except those with respect to privileges.

(b) Relevancy conditioned on fact

When the relevancy of evidence depends upon the fulfillment of a condition of fact, the court shall admit it upon, or subject to, the introduction of evidence sufficient to support a finding of the fulfillment of the condition.

(c) Hearing of jury

Hearings on the admissibility of confessions shall in all cases be conducted out of the hearing of the jury. Hearings on other preliminary matters shall be so conducted when the interests of justice require, or when an accused is a witness and so request.

(d) Testimony by accused

The accused does not, by testifying upon a preliminary matter, become subject to cross-examination as to other issues in the case.

(e) Weight and credibility

This rule does not limit the right of a party to introduce before the jury evidence relevant to weight or credibility.

Article VII : Opinions and Expert Testimony

Rule 701. Opinion Testimony by Lay Witnesses

If the witness is not testifying as an expert, the witness' testimony in the form of opinions or inferences is limited to those opinions or inferences which are (a) rationally based on the perception of the witness and (b) helpful to a clear understanding of the witness' testimony or the determination of a fact in issue.

Rule 702. Testimony by Experts

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

Rule 703. Bases of Opinion Testimony by Experts

The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to the expert at or before the hearing. If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence.

Article IX: Authentication and Identification

Rule 901. Requirement of Authentication or Identification

(a) General provision

The requirement of authentication or identification as a condition precedent to admissibility is satisfied by evidence sufficient to support a finding that the matter in question is what its proponent claims.

(b) Illustrations

By way of illustration only, and not by way of limitation, the following are examples of authentication or identification conforming with the requirements of this rule:

(1) Testimony of witness with knowledge

Testimony that a matter is what it is claimed to be.

(2) Nonexpert opinion on handwriting

Nonexpert opinion as to the genuineness of handwriting, based upon familiarity not acquired for purposes of the litigation.

(3) Comparison by trier or expert witness

Comparison by the trier of fact or by expert witnesses with specimens which have been authenticated.

(4) Distinctive characteristics and the like

Appearance, contents, substance, internal patterns, or other distinctive characteristics, taken in conjunction with circumstances.

(5) Voice identification

Identification of a voice, whether heard firsthand or through mechanical or electronic transmission or recording, by opinion based upon hearing the voice at any time under circumstances connecting it with the alleged speaker.

(6) Telephone conversations

Telephone conversations, by evidence that a call was made to the number assigned at the time by the telephone company to a particular person or business, if (A) in the case of a person, circumstances, including self-identification, show the person answering to be the one called, or (B) in the case of a business, the call was made to a place of business and the conversation related to business reasonably transacted over the telephone.

(7) Public records or reports

Evidence that a writing authorized by law to be recorded or filed and in fact recorded or filed in a public office, or a purported public record, report, statement, or data compilation, in any form, is from the public office where items of this nature are kept.

(8) Ancient documents or data compilation

Evidence that a document or data compilation, in any form, (A) is in such condition as to create no suspicion concerning its authenticity, (B) was in a place where it, if authentic, would likely be, and (C) has been in existence 20 years or more at the time it is offered.

(9) Process or system

Evidence describing a process or system used to produce a result and showing that the process or system produces an accurate result.

(10) Methods provided by statute or rule

Any method of authentication or identification provided by Act of Congress or by other rules prescribed by the Supreme Court pursuant to statutory authority.

Annexe V. Code de procédure de l'*International Association for Forensic Phonetics* (IAFP)

L'*International Association for Forensic Phonetics* (IAFP) a été formellement établie après le troisième séminaire annuel sur la phonétique forensique à York, Angleterre, du 24 au 27 juin 1991.

Les buts de l'association devraient être :

1. D'entretenir la recherche et de prévoir un forum pour l'échange d'idées et d'informations sur la pratique, le développement et la recherche en phonétique forensique.
2. D'établir par écrit et de renforcer les standards de conduite professionnelle et de procédure pour ceux qui sont engagés dans la pratique de l'expertise en phonétique forensique.

Code de procédure

1. L'analyse forensique de la parole devrait être prise en charge seulement par ceux qui ont un entraînement et des qualifications en phonétique – sciences de la parole.
2. Les membres devraient toujours agir avec intégrité, équité et impartialité.
3. Les membres devraient décliner leur affiliation à l'IAFP dans leurs rapports et lors de leurs témoignages en cour.
4. Les membres devraient mentionner clairement les limitations de l'analyse forensique de la parole dans leurs rapports et en cour.
5. Les membres devraient mentionner clairement leur degré de certitude dans leur conclusion et donner une indication de l'endroit où elle se situe dans l'échelle des conclusions qu'ils sont prêts à donner.
6. (a) Les membres devraient approcher avec la plus grande prudence l'analyse forensique d'échantillons de parole énoncés dans une autre langue que leur langue maternelle.
6. (b) Les membres devraient approcher avec la plus grande prudence l'analyse forensique d'échantillons de parole énoncés dans plusieurs langues.
7. Les membres devraient préciser dans leurs rapports les méthodes d'analyse sur lesquelles leur conclusion est basée.
8. Les membres, en faisant leur analyse, devraient tenir compte des méthodes disponibles et de leur opportunité pour l'analyse des échantillons.
9. Les membres ne devraient pas effectuer de profils psychologiques des locuteurs, ni se prononcer sur leur sincérité.

ANNEXE VI. BASE DE DONNEES POLYPHONE IPSC

A.VI.1. Date des sessions d'enregistrement

A.VI.1.a. Enregistrements des modèles

Locutrice	Session Polyphone cellulaire	Session Polyphone 1	Session Polyphone 2	Session Polyphone 3	Session Polyphone 4	Session Polyphone 5	Session Comparaison
00	J + 0	J + 0	J + 8	J + 26	J + 28	J + 28	J + 0
01	J + 0	J + 0	J + 12	J + 16	J + 26	J + 33	J + 0
04	J + 0	J + 0	J + 22	J + 29	J + 35	J + 42	J + 0
05	J + 0	J + 0	J + 8	J + 11	J + 20	J + 45	J + 0
06	J + 0	J + 0	J + 19	J + 25	J + 36	J + 64	J + 0
07	J + 0	J + 0	J + 10	J + 18	J + 26	J + 32	J + 0
08	J + 0	J + 0	J + 7	J + 15	J + 22	J + 34	J + 0
09	J + 0	J + 0	J + 8	J + 22	J + 25	J + 32	J + 0
32	J + 0	J + 0	J + 14	J + 28	J + 35	J + 42	J + 0
33	J + 0	J + 0	J + 12	J + 17	J + 32	J + 38	J + 0
44	J + 0	J + 0	J + 14	J + 28	J + 36	J + 42	J + 0
49	J + 0	J + 0	J + 10	J + 15	J + 34	J + 38	J + 0
54	J + 63	J + 0	J + 9	J + 18	J + 25	J + 34	J + 0
55	J + 32	J + 0	J + 10	J + 21	J + 29	J + 32	J + 0
58	J + 22	J + 0	J + 21	J + 27	J + 29	J + 43	J + 0
59	J + 0	J + 0	J + 8	J + 15	J + 20	J + 29	J + 0

Locuteur	Session Polyphone cellulaire	Session Polyphone 1	Session Polyphone 2	Session Polyphone 3	Session Polyphone 4	Session Polyphone 5	Session Comparaison
10	J + 0	J + 0	J + 9	J + 15	J + 52	J + 53	J + 0
11	J + 0	J + 13	J + 18	J + 27	J + 28	J + 40	J + 0
12	J + 0	J + 0	J + 27	J + 33	J + 45	J + 59	J + 0
13	J + 0	J + 0	J + 10	J + 14	J + 28	J + 37	J + 0
14	J + 0	J + 0	J + 7	J + 16	J + 36	J + 58	J + 0
15	J + 0	J + 0	J + 10	J + 18	J + 27	J + 32	J + 0
16	J + 0	J + 0	J + 9	J + 18	J + 21	J + 29	J + 0
17	J + 0	J + 0	J + 9	J + 14	J + 18	J + 21	J + 0
18	J + 0	J + 0	J + 5	J + 12	J + 15	J + 22	J + 0
19	J + 0	J + 0	J + 25	J + 25	J + 42	J + 60	J + 0
20	J + 0	J + 0	J + 7	J + 13	J + 23	J + 28	J + 0
22	J + 0	J + 0	J + 11	J + 32	J + 58	J + 93	J + 0
39	J + 0	J + 0	J + 3	J + 6	J + 8	J + 13	J + 0
40	J + 0	J + 0	J + 3	J + 13	J + 16	J + 23	J + 0
41	J + 0	J + 0	J + 9	J + 14	J + 30	J + 35	J + 0
56	J + 0	J + 0	J + 14	J + 18	J + 27	J + 33	J + 0

A.VI.1.b. Enregistrements de test

Locutrice	Test cellulaire	Test 1	Test 2	Test 3	Test 4	Test 5	Message anonyme 1	Message anonyme 2
00	J + 0	J + 0	J + 8	J + 26	J + 28	J + 28	J + 32	J + 32
01	J + 0	J + 0	J + 12	J + 16	J + 26	J + 33	J + 54	J + 55
04	J + 0	J + 0	J + 22	J + 29	J + 35	J + 42	J + 83	J + 83
05	J + 0	J + 0	J + 8	J + 11	J + 20	J + 45	J + 59	J + 95
06	J + 0	J + 0	J + 19	J + 25	J + 36	J + 64	J + 21	J + 25
07	J + 0	J + 0	J + 10	J + 18	J + 26	J + 32	J + 26	J + 32
08	J + 0	J + 0	J + 7	J + 15	J + 22	J + 34	J + 7	J + 15
09	J + 0	J + 0	J + 8	J + 22	J + 25	J + 32	J + 22	J + 25
32	J + 0	J + 0	J + 14	J + 28	J + 35	J + 42	J + 14	J + 28
33	J + 0	J + 0	J + 12	J + 17	J + 32	J + 38	J + 45	J + 53
44	J + 0	J + 0	J + 14	J + 28	J + 36	J + 42	J + 14	J + 14
49	J + 0	J + 0	J + 10	J + 15	J + 34	J + 38	J + 11	J + 25
54	J + 63	J + 0	J + 9	J + 18	J + 25	J + 34	J + 18	J + 67
55	J + 32	J + 0	J + 10	J + 21	J + 29	J + 32	J + 74	J + 74
58	J + 22	J + 0	J + 21	J + 27	J + 29	J + 43	J + 21	J + 50
59	J + 0	J + 0	J + 8	J + 15	J + 20	J + 29	J + 68	J + 69

Locuteur	Test cellulaire	Test 1	Test 2	Test 3	Test 4	Test 5	Message anonyme 1	Message anonyme 2
10	J + 0	J + 0	J + 9	J + 15	J + 52	J + 53	J + 91	J + 105
11	J + 0	J + 13	J + 18	J + 27	J + 28	J + 40	J + 0	J + 18
12	J + 0	J + 0	J + 27	J + 33	J + 45	J + 59	J + 37	J + 39
13	J + 0	J + 0	J + 10	J + 14	J + 28	J + 37	J + 30	J + 43
14	J + 0	J + 0	J + 7	J + 16	J + 36	J + 58	J + 16	J + 36
15	J + 0	J + 0	J + 10	J + 18	J + 27	J + 32	J + 30	J + 30
16	J + 0	J + 0	J + 9	J + 18	J + 21	J + 29	J + 17	J + 21
17	J + 0	J + 0	J + 9	J + 14	J + 18	J + 21	J + 9	J + 37
18	J + 0	J + 0	J + 5	J + 12	J + 15	J + 22	J + 41	J + 41
19	J + 0	J + 0	J + 25	J + 25	J + 42	J + 60	J + 60	J + 60
20	J + 0	J + 0	J + 7	J + 13	J + 23	J + 28	J + 23	J + 86
22	J + 0	J + 0	J + 11	J + 32	J + 58	J + 93	J + 58	J + 58
39	J + 0	J + 0	J + 3	J + 6	J + 8	J + 13	J + 13	J + 17
40	J + 0	J + 0	J + 3	J + 13	J + 16	J + 23	J + 14	J + 22
41	J + 0	J + 0	J + 9	J + 14	J + 30	J + 35	J + 31	J + 37
56	J + 0	J + 0	J + 14	J + 18	J + 27	J + 33	J + 27	J + 27

Locuteur	Session Polyphone cellulaire	Session Polyphone 1	Session Polyphone 2	Session Polyphone 3	Session Polyphone 4	Session Polyphone 5	Session Comparaison
10*	GSM (70)	RTPC (44)	RTPC (44)	RTPC (44)	RTPC (44)	RTPC (44)	RTPC (44)
11	GSM (70)	RTPC (38)	RTPC (38)	RTPC (38)	DECT (38)	RTPC (38)	RTPC (38)
12	GSM (70)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)
13*	GSM (70)	DECT (26)	DECT (26)	DECT (26)	DECT (26)	DECT (26)	DECT (26)
14	GSM (70)	RTPC (47)	DECT (47)	DECT (47)	DECT (47)	DECT (47)	RTPC (47)
15*	GSM (70)	DECT (67)	DECT (67)	DECT (67)	DECT (67)	DECT (67)	DECT (67)
16*	GSM (70)	DECT (63)	DECT (63)	DECT (63)	DECT (63)	DECT (63)	DECT (63)
17	GSM (70)	DECT (63)	DECT (63)	DECT (63)	DECT (63)	DECT (44)	DECT (63)
18*	GSM (70)	RTPC (40)	RTPC (40)	RTPC (40)	RTPC (40)	RTPC (40)	RTPC (40)
19*	GSM (70)	DECT (37)	DECT (37)	DECT (37)	DECT (37)	DECT (37)	DECT (37)
20*	GSM (70)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)
22*	GSM (70)	RTPC (25)	RTPC (25)	RTPC (25)	RTPC (25)	RTPC (25)	RTPC (25)
39*	GSM (70)	DECT (96)	DECT (96)	DECT (96)	DECT (96)	DECT (96)	DECT (96)
40	GSM (70)	RTPC (09)	RTPC (09)	RTPC (30)	RTPC (30)	RTPC (30)	RTPC (09)
41	GSM (70)	RTPC (28)	RTPC (30)	RTPC (30)	RTPC (30)	RTPC (30)	RTPC (28)
56	GSM (70)	RTPC (44)	RTPC (44)	GSM (77)	RTPC (21)	RTPC (44)	RTPC (44)

A.VI.2.b. Enregistrements de comparaison

Locutrice	Session Comparaison	Locutrice	Session Comparaison	Locuteur	Session Comparaison	Locuteur	Session Comparaison
00	DECT (17)	32	RTPC (32)	10	RTPC (44)	18	RTPC (40)
01	DECT (17)	33	DECT (01)	11	DECT (38)	19	DECT (37)
04	RTPC (28)	44	RTPC (32)	12	RTPC (16)	20	RTPC (16)
05	RTPC (08)	49	RTPC (28)	13	DECT (26)	22	RTPC (25)
06	RTPC (11)	54	RTPC (38)	14	RTPC (47)	39	DECT (96)
07	DECT (67)	55	RTPC (38)	15	DECT (67)	40	RTPC (09)
08	DECT (81)	58	RTPC (75)	16	DECT (63)	41	RTPC (28)
09	DECT (01)	59	RTPC (09)	17	DECT (63)	56	RTPC (44)

A.VI.2.c. Enregistrements de test

Locutrice	Test cellulaire	Test 1	Test 2	Test 3	Test 4	Test 5	Message anonyme 1	Message anonyme 2
00	GSM (70)	RTPC (17)	RTPC (17)	DECT (17)	DECT (17)	DECT (17)	RTPC (17)	RTPC (17)
01	GSM (70)	RTPC (17)	RTPC (17)	RTPC (17)	RTPC (17)	RTPC (17)	DECT (17)	DECT (17)
04	GSM (70)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)
05	GSM (70)	RTPC (08)	RTPC (08)	RTPC (16)	RTPC (08)	RTPC (08)	RTPC (08)	RTPC (08)
06	GSM (70)	RTPC (11)	RTPC (11)	RTPC (11)	RTPC (11)	RTPC (11)	RTPC (11)	RTPC (11)
07	GSM (70)	DECT (67)	DECT (67)	DECT (67)	RTPC (67)	DECT (67)	RTPC (67)	DECT (67)
08	GSM (70)	DECT (81)	RTPC (81)	RTPC (81)	RTPC (81)	RTPC (81)	DECT (81)	RTPC (81)
09	GSM (70)	DECT (01)	DECT (00)	DECT (00)	DECT (00)	DECT (00)	DECT (00)	DECT (00)
32	GSM (70)	RTPC (32)	RTPC (32)	RTPC (32)	RTPC (32)	RTPC (32)	RTPC (32)	RTPC (32)
33	GSM (70)	DECT (01)	DECT (01)	DECT (01)	DECT (01)	DECT (01)	DECT (01)	DECT (01)
44	GSM (70)	RTPC (32)	RTPC (32)	RTPC (32)	RTPC (32)	RTPC (32)	RTPC (32)	RTPC (32)
49	GSM (70)	RTPC (28)	RTPC (16)	RTPC (16)	RTPC (32)	RTPC (21)	RTPC (16)	GSM (70)
54	GSM (70)	RTPC (38)	RTPC (38)	RTPC (38)	RTPC (38)	RTPC (38)	RTPC (38)	RTPC (38)
55	GSM (70)	RTPC (38)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)
58	GSM (70)	RTPC (75)	RTPC (75)	RTPC (75)	RTPC (75)	RTPC (09)	RTPC (75)	RTPC (75)
59	GSM (70)	RTPC (09)	RTPC (09)	RTPC (09)	RTPC (09)	RTPC (09)	RTPC (09)	RTPC (09)

Locuteur	Test cellulaire	Test 1	Test 2	Test 3	Test 4	Test 5	Message anonyme 1	Message anonyme 2
10	GSM (70)	RTPC (44)	RTPC (44)	RTPC (44)	RTPC (44)	RTPC (44)	RTPC (44)	RTPC (44)
11	GSM (70)	RTPC (38)	RTPC (38)	RTPC (38)	DECT (38)	RTPC (38)	GSM (70)	RTPC (38)
12	GSM (70)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)	RTPC (28)
13	GSM (70)	DECT (26)	DECT (26)	DECT (26)	DECT (26)	DECT (26)	DECT (26)	DECT (26)
14	GSM (70)	RTPC (47)	DECT (47)	DECT (47)	DECT (47)	DECT (47)	RTPC (47)	RTPC (47)
15	GSM (70)	DECT (67)	DECT (67)	DECT (67)	DECT (67)	DECT (67)	DECT (67)	DECT (67)
16	GSM (70)	DECT (63)	DECT (63)	DECT (63)	DECT (63)	DECT (63)	DECT (63)	DECT (63)
17	GSM (70)	DECT (63)	DECT (63)	DECT (63)	DECT (63)	DECT (44)	DECT (63)	DECT (63)
18	GSM (70)	RTPC (40)	RTPC (40)	RTPC (40)	RTPC (40)	RTPC (40)	RTPC (40)	RTPC (40)
19	GSM (70)	DECT (37)	DECT (37)	DECT (37)	DECT (37)	DECT (37)	DECT (37)	DECT (37)
20	GSM (70)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)	RTPC (16)
22	GSM (70)	RTPC (25)	RTPC (25)	RTPC (25)	RTPC (25)	RTPC (25)	RTPC (25)	RTPC (25)
39	GSM (70)	DECT (96)	DECT (96)	DECT (96)	DECT (96)	DECT (96)	DECT (96)	DECT (96)
40	GSM (70)	RTPC (09)	RTPC (09)	RTPC (30)	RTPC (30)	RTPC (30)	RTPC (30)	RTPC (30)
41	GSM (70)	RTPC (28)	RTPC (30)	RTPC (30)	RTPC (30)	RTPC (30)	RTPC (28)	RTPC (30)
56	GSM (70)	RTPC (44)	RTPC (44)	GSM (77)	RTPC (21)	RTPC (44)	RTPC (21)	RTPC (21)

A.VI.3. Composition des enregistrements

Enregistrements de comparaison

(-c) indique la session d'enregistrement de comparaison d'une cinquantaine de diapositives

(-ad) indique la simulation d'un message anonyme avec un crayon dans la bouche

(-an) indique la simulation d'un message anonyme avec la voix normale

(-ld) indique un texte lu avec un crayon dans la bouche

(-d1) et (-d2) indiquent la simulation de dialogues

(-s) indique de la parole spontanée

Le nombre placé en dernier indique la durée en secondes

A.VI.3.a. Enregistrements de comparaison des locutrices L00 - L09

Locutrice 00	Locutrice 01	Locutrice 04	Locutrice 05	Locutrice 06	Locutrice 07	Locutrice 08	Locutrice 09
Simulation de messages anonymes							
L00-c-ad-08.3	L01-c-ad-12.3	L04-c-ad-07.7	L05-c-ad-09.2	L06-c-ad-09.9	L07-c-ad-12.0		L09-c-ad-10.3
L00-c-an-07.7	L01-c-an-13.1	L04-c-an-08.9	L05-c-an-07.8	L06-c-an-09.9	L07-c-ad-12.1		L09-c-an-10.0
					L07-c-an-10.7		
Simulation de dialogues							
L00-c-d1-1.4	L01-c-d1-01.3	L04-c-d1-01.4	L05-c-d1-01.6	L06-c-d1-01.6	L07-c-d1-01.9	L08-c-d1-01.3	L09-c-d1-01.4
L00-c-d1-1.6	L01-c-d1-02.5	L04-c-d1-01.6	L05-c-d1-01.7	L06-c-d1-02.0	L07-c-d1-02.0	L08-c-d1-02.1	L09-c-d1-01.8
L00-c-d1-2.5	L01-c-d1-03.6	L04-c-d1-02.0	L05-c-d1-02.1	L06-c-d1-02.1	L07-c-d1-02.2	L08-c-d1-02.2	L09-c-d1-01.9
L00-c-d1-2.6	L01-c-d1-1.5	L04-c-d1-03.4	L05-c-d1-02.3	L06-c-d1-02.7	L07-c-d1-02.4	L08-c-d1-02.3	L09-c-d1-05.8
L00-c-d2-1.6	L01-c-d2-03.1	L04-c-d2-01.2	L05-c-d2-00.8	L06-c-d2-01.4	L07-c-d2-01.4	L08-c-d2-01.2	L09-c-d2-01.9
L00-c-d2-2.2	L01-c-d2-04.0	L04-c-d2-02.2	L05-c-d2-01.5	L06-c-d2-02.5	L07-c-d2-02.4	L08-c-d2-02.1	L09-c-d2-02.0
L00-c-d2-3.0	L01-c-d2-07.0	L04-c-d2-03.2	L05-c-d2-02.3	L06-c-d2-02.6	L07-c-d2-05.5	L08-c-d2-02.8	L09-c-d2-02.6
L00-c-d2-5.1		L04-c-d2-05.9	L05-c-d2-04.9			L08-c-d2-05.2	L09-c-d2-02.7
Lecture déguisée							
L00-c-ld-07.3	L01-c-ld-11.3	L04-c-ld-04.2	L05-c-ld-02.7	L06-c-ld-06.4	L07-c-ld-04.9	L08-c-ld-04.2	L09-c-ld-04.6
L00-c-ld-10.7	L01-c-ld-14.6	L04-c-ld-06.4	L05-c-ld-03.5	L06-c-ld-11.3	L07-c-ld-07.1	L08-c-ld-06.6	L09-c-ld-08.5
L00-c-ld-11.0	L01-c-ld-14.8	L04-c-ld-07.0	L05-c-ld-03.7	L06-c-ld-12.0	L07-c-ld-09.1	L08-c-ld-07.2	L09-c-ld-12.5
L00-c-ld-11.5	L01-c-ld-16.3	L04-c-ld-10.2	L05-c-ld-06.0	L06-c-ld-12.5	L07-c-ld-09.5	L08-c-ld-10.8	L09-c-ld-12.6
	L01-c-ld-11.9	L04-c-ld-10.8	L05-c-ld-06.3		L07-c-ld-13.2	L08-c-ld-11.2	
			L05-c-ld-06.9				
			L05-c-ld-07.0				
			L05-c-ld-10.6				
Parole spontanée							
L00-c-s-01.5	L01-c-s-07.0	L04-c-s-01.0	L05-c-s-02.4	L06-c-s-02.1	L07-c-s-01.5	L08-c-s-01.5	L09-c-s-07.6
L00-c-s-04.3	L01-c-s-08.6	L04-c-s-06.9	L05-c-s-05.6	L06-c-s-02.2	L07-c-s-02.5	L08-c-s-07.0	L09-c-s-08.8
L00-c-s-06.4	L01-c-s-10.2	L04-c-s-08.7	L05-c-s-07.2	L06-c-s-03.1	L07-c-s-03.1	L08-c-s-07.5	L09-c-s-09.1
L00-c-s-06.7	L01-c-s-10.3	L04-c-s-08.8	L05-c-s-07.8	L06-c-s-06.6	L07-c-s-03.2	L08-c-s-08.2	L09-c-s-10.1
L00-c-s-07.4	L01-c-s-10.9	L04-c-s-09.0	L05-c-s-08.4	L06-c-s-06.9	L07-c-s-03.6	L08-c-s-08.9	L09-c-s-10.8
L00-c-s-07.8	L01-c-s-11.5	L04-c-s-09.3	L05-c-s-08.6	L06-c-s-07.0	L07-c-s-03.7	L08-c-s-10.3	L09-c-s-11.2
L00-c-s-07.9	L01-c-s-12.5	L04-c-s-09.4	L05-c-s-10.2	L06-c-s-07.4	L07-c-s-05.1	L08-c-s-10.7	L09-c-s-11.3
L00-c-s-08.5	L01-c-s-16.4	L04-c-s-10.8	L05-c-s-11.0	L06-c-s-08.9	L07-c-s-05.8	L08-c-s-11.1	L09-c-s-11.9
L00-c-s-08.9	L01-c-s-18.0	L04-c-s-11.0	L05-c-s-11.2	L06-c-s-10.1	L07-c-s-06.0	L08-c-s-12.6	L09-c-s-13.9
L00-c-s-09.3	L01-c-s-18.7	L04-c-s-11.3	L05-c-s-11.3	L06-c-s-10.5	L07-c-s-06.3	L08-c-s-12.8	L09-c-s-14.5

Locutrice 00	Locutrice 01	Locutrice 04	Locutrice 05	Locutrice 06	Locutrice 07	Locutrice 08	Locutrice 09
L00-c-s-10.4	L01-c-s-20.1	L04-c-s-11.4	L05-c-s-11.6	L06-c-s-10.7	L07-c-s-06.4	L08-c-s-13.3	L09-c-s-16.4
L00-c-s-10.5	L01-c-s-23.3	L04-c-s-12.1	L05-c-s-13.8	L06-c-s-11.0	L07-c-s-06.8	L08-c-s-14.2	L09-c-s-17.6
L00-c-s-11.5	L01-c-s-23.4	L04-c-s-12.3	L05-c-s-15.6	L06-c-s-13.3	L07-c-s-07.7	L08-c-s-14.7	L09-c-s-17.9
L00-c-s-11.6		L04-c-s-12.4	L05-c-s-15.7	L06-c-s-13.9	L07-c-s-07.8	L08-c-s-16.2	L09-c-s-18.5
L00-c-s-11.7		L04-c-s-12.5	L05-c-s-18.0	L06-c-s-14.5	L07-c-s-07.9	L08-c-s-17.0	L09-c-s-18.6
L00-c-s-11.8		L04-c-s-12.6		L06-c-s-15.9	L07-c-s-08.4	L08-c-s-17.1	L09-c-s-19.3
L00-c-s-11.9		L04-c-s-13.3		L06-c-s-16.3	L07-c-s-09.1	L08-c-s-17.9	L09-c-s-20.1
L00-c-s-12.0		L04-c-s-14.1		L06-c-s-22.4	L07-c-s-09.2	L08-c-s-18.0	L09-c-s-22.5
L00-c-s-12.3		L04-c-s-15.5			L07-c-s-09.3	L08-c-s-19.6	L09-c-s-28.9
L00-c-s-12.6					L07-c-s-09.4	L08-c-s-19.7	
L00-c-s-13.9					L07-c-s-09.5	L08-c-s-22.7	
L00-c-s-14.2					L07-c-s-09.7		
L00-c-s-15.0					L07-c-s-09.8		
L00-c-s-15.4					L07-c-s-09.9		
L00-c-s-18.0					L07-c-s-10.1		
L00-c-s-19.5					L07-c-s-11.6		
L00-c-s-24.7					L07-c-s-11.7		
					L07-c-s-11.9		
					L07-c-s-12.4		
					L07-c-s-12.5		
					L07-c-s-12.6		
					L07-c-s-12.8		
					L07-c-s-14.4		
					L07-c-s-15.0		
					L07-c-s-15.6		
					L07-c-s-17.2		
					L07-c-s-17.3		
					L07-c-s-17.4		
					L07-c-s-24.8		

A.VI.3.b. Enregistrements de comparaison des locutrices L32 - L59

Locutrice 32	Locutrice 33	Locutrice 44	Locutrice 49	Locutrice 54	Locutrice 55	Locutrice 58	Locutrice 59
Simulation de messages anonymes							
L32-c-ad-11.8	L33-c-ad-07.6	L44-c-an-11.6	L49-c-ad-07.5	L54-c-an-09.5	L55-c-ad-07.4	L58-c-ad-13.1	L59-c-ad-10.4
L32-c-an-11.7	L33-c-an-07.4		L49-c-an-10.6		L55-c-an-08.3	L58-c-an-09.6	L59-c-ad-13.5
							L59-c-an-08.5
Simulation de dialogues							
L32-c-d1-02.6	L33-c-d1-01.2	L44-c-d1-02.2	L49-c-d1-00.9	L54-c-d1-01.7	L55-c-d1-01.3	L58-c-d1-01.6	L59-c-d1-01.6
L32-c-d1-02.9	L33-c-d1-01.5	L44-c-d1-03.7	L49-c-d1-01.3	L54-c-d1-01.8	L55-c-d1-01.4	L58-c-d1-02.0	L59-c-d1-02.0
L32-c-d1-03.1	L33-c-d1-01.8	L44-c-d1-04.4	L49-c-d1-01.7	L54-c-d1-01.9	L55-c-d1-01.6	L58-c-d1-02.7	L59-c-d1-02.5
L32-c-d1-03.2	L33-c-d1-01.9	L44-c-d1-06.8	L49-c-d1-02.7	L54-c-d1-02.5	L55-c-d1-02.6	L58-c-d1-03.2	L59-c-d1-02.8
L32-c-d2-02.8	L33-c-d2-01.1	L44-c-d2-01.8	L49-c-d2-01.4	L54-c-d2-01.0	L55-c-d2-00.9	L58-c-d2-01.0	L59-c-d2-01.3
L32-c-d2-04.4	L33-c-d2-02.1	L44-c-d2-03.2	L49-c-d2-02.4	L54-c-d2-02.3	L55-c-d2-01.8	L58-c-d2-01.9	L59-c-d2-01.8
L32-c-d2-07.0	L33-c-d2-02.6	L44-c-d2-04.4	L49-c-d2-03.8	L54-c-d2-02.6	L55-c-d2-02.1	L58-c-d2-02.3	L59-c-d2-04.4
	L33-c-d2-03.7	L44-c-d2-09.6		L54-c-d2-06.0	L55-c-d2-06.8	L58-c-d2-05.5	L59-c-d2-06.4
Lecture déguisée							
L32-c-ld-06.2	L33-c-ld-03.2	L44-c-ld-10.0	L49-c-ld-02.6	L54-c-ld-03.7	L55-c-ld-03.6	L58-c-ld-03.8	L59-c-ld-04.7
L32-c-ld-12.8	L33-c-ld-05.1	L44-c-ld-11.5	L49-c-ld-06.6	L54-c-ld-05.3	L55-c-ld-06.5	L58-c-ld-04.4	L59-c-ld-07.4
L32-c-ld-15.2	L33-c-ld-06.9	L44-c-ld-13.4	L49-c-ld-06.9	L54-c-ld-06.9	L55-c-ld-06.6	L58-c-ld-07.5	L59-c-ld-08.5
L32-c-ld-15.9	L33-c-ld-08.5	L44-c-ld-13.7	L49-c-ld-08.4	L54-c-ld-09.9	L55-c-ld-11.0	L58-c-ld-08.9	L59-c-ld-09.4
L32-c-ld-22.1	L33-c-ld-08.7	L44-c-ld-22.3	L49-c-ld-09.5	L54-c-ld-10.5	L55-c-ld-13.0	L58-c-ld-24.2	L59-c-ld-12.4
		L44-c-ld-26.6					L59-c-ld-16.2
							L59-c-ld-16.8
Parole spontanée							
L32-c-s-02.2	L33-c-s-01.1	L44-c-s-01.9	L49-c-s-02.2	L54-c-s-00.8	L55-c-s-02.9	L58-c-s-01.5	L59-c-s-04.5
L32-c-s-03.5	L33-c-s-02.0	L44-c-s-03.6	L49-c-s-02.3	L54-c-s-02.1	L55-c-s-04.0	L58-c-s-01.6	L59-c-s-06.0
L32-c-s-03.6	L33-c-s-02.2	L44-c-s-03.8	L49-c-s-03.2	L54-c-s-02.9	L55-c-s-04.1	L58-c-s-05.9	L59-c-s-06.5
L32-c-s-03.7	L33-c-s-02.5	L44-c-s-04.1	L49-c-s-03.3	L54-c-s-03.1	L55-c-s-04.2	L58-c-s-06.1	L59-c-s-06.6
L32-c-s-04.1	L33-c-s-02.7	L44-c-s-05.7	L49-c-s-04.5	L54-c-s-04.3	L55-c-s-04.9	L58-c-s-08.2	L59-c-s-07.5
L32-c-s-04.2	L33-c-s-02.8	L44-c-s-05.8	L49-c-s-05.2	L54-c-s-04.8	L55-c-s-05.0	L58-c-s-08.5	L59-c-s-07.9
L32-c-s-04.3	L33-c-s-02.9	L44-c-s-05.9	L49-c-s-06.0	L54-c-s-04.9	L55-c-s-05.1	L58-c-s-08.9	L59-c-s-08.1
L32-c-s-04.5	L33-c-s-03.0	L44-c-s-06.0	L49-c-s-08.6	L54-c-s-05.1	L55-c-s-05.2	L58-c-s-09.4	L59-c-s-08.3
L32-c-s-04.6	L33-c-s-03.1	L44-c-s-07.0	L49-c-s-09.0	L54-c-s-07.0	L55-c-s-05.5	L58-c-s-10.2	L59-c-s-10.4
L32-c-s-04.7	L33-c-s-03.2	L44-c-s-08.1	L49-c-s-09.5	L54-c-s-07.4	L55-c-s-05.6	L58-c-s-10.6	L59-c-s-11.8
L32-c-s-05.8	L33-c-s-03.3	L44-c-s-08.3	L49-c-s-09.6	L54-c-s-08.0	L55-c-s-06.5	L58-c-s-10.7	L59-c-s-11.9
L32-c-s-06.2	L33-c-s-03.7	L44-c-s-08.4	L49-c-s-09.8	L54-c-s-09.1	L55-c-s-07.4	L58-c-s-11.3	L59-c-s-112.0
L32-c-s-06.9	L33-c-s-03.9	L44-c-s-09.7	L49-c-s-10.2	L54-c-s-09.2	L55-c-s-07.5	L58-c-s-11.5	L59-c-s-18.7
L32-c-s-09.0	L33-c-s-05.0	L44-c-s-09.8	L49-c-s-12.3	L54-c-s-10.2	L55-c-s-07.9	L58-c-s-13.3	L59-c-s-19.5
L32-c-s-09.9	L33-c-s-05.8		L49-c-s-12.8	L54-c-s-10.8	L55-c-s-08.1	L58-c-s-13.9	L59-c-s-20.6
L32-c-s-13.9	L33-c-s-05.9		L49-c-s-13.4	L54-c-s-13.8		L58-c-s-14.1	
	L33-c-s-06.2		L49-c-s-14.5			L58-c-s-15.4	
			L49-c-s-14.9			L58-c-s-16.7	
			L49-c-s-15.0			L58-c-s-17.5	
			L49-c-s-15.6			L58-c-s-17.7	
			L49-c-s-19.1			L58-c-s-18.7	
			L49-c-s-19.7			L58-c-s-20.2	
			L49-c-s-20.7			L58-c-s-20.4	
			L49-c-s-21.0			L58-c-s-28.6	
			L49-c-s-21.1				
			L49-c-s-21.4				
			L49-c-s-22.0				
			L49-c-s-23.1				
			L49-c-s-23.8				

A.VI.3.c. Enregistrements de comparaison des locuteurs L10 - L17

Locuteur 10	Locuteur 11	Locuteur 12	Locuteur 13	Locuteur 14	Locuteur 15	Locuteur 16	Locuteur 17
Simulation de messages anonymes							
L10-c-ad-12.1	L11-c-ad-05.2	L12-c-ad-08.1	L13-c-ad-07.9	L14-c-ad-07.5	L15-c-an-07.7	L16-c-ad-07.7	L17-c-ad-09.6
L10-c-an-10.6	L11-c-ad-07.2	L12-c-an-06.8	L13-c-an-08.2	L14-c-an-07.0		L16-c-an-06.9	L17-c-an-09.4
	L11-c-an-06.6						
Simulation de dialogues							
L10-c-d1-01.5	L11-c-d1-01.3	L12-c-d1-00.7	L13-c-d1-01.5	L14-c-d1-01.3	L15-c-d1-01.3	L16-c-d1-01.3	L17-c-d1-01.6
L10-c-d1-01.9	L11-c-d1-01.6	L12-c-d1-01.0	L13-c-d1-01.6	L14-c-d1-01.4	L15-c-d1-01.4	L16-c-d1-01.4	L17-c-d1-01.7
L10-c-d1-02.2	L11-c-d1-01.8	L12-c-d1-01.6	L13-c-d1-02.3	L14-c-d1-01.7	L15-c-d1-01.7	L16-c-d1-01.7	L17-c-d1-02.6
L10-c-d1-02.4	L11-c-d1-02.1	L12-c-d1-01.8	L13-c-d1-03.0	L14-c-d1-02.5	L15-c-d1-02.6	L16-c-d1-02.1	L17-c-d1-02.7
L10-c-d2-01.7	L11-c-d2-00.9	L12-c-d2-00.9	L13-c-d1-06.1	L14-c-d2-01.2	L15-c-d2-01.1	L16-c-d2-01.0	L17-c-d2-01.0
L10-c-d2-01.8	L11-c-d2-01.4	L12-c-d2-01.2	L13-c-d2-00.9	L14-c-d2-02.0	L15-c-d2-01.9	L16-c-d2-01.9	L17-c-d2-01.9
L10-c-d2-03.5	L11-c-d2-03.8	L12-c-d2-02.0	L13-c-d2-02.6	L14-c-d2-02.8	L15-c-d2-04.7	L16-c-d2-02.5	L17-c-d2-02.3
L10-c-d2-05.0		L12-c-d2-04.6	L13-c-d2-02.7	L14-c-d2-04.9	L15-c-d2-04.9	L16-c-d2-04.5	L17-c-d2-04.6
Lecture déguisée							
L10-c-ld-05.0	L11-c-ld-02.5	L12-c-ld-09.6	L13-c-ld-04.7	L14-c-ld-06.0	L15-c-ld-04.0	L16-c-ld-03.3	L17-c-ld-02.5
L10-c-ld-07.7	L11-c-ld-03.4	L12-c-ld-11.5	L13-c-ld-06.6	L14-c-ld-09.4	L15-c-ld-05.1	L16-c-ld-05.5	L17-c-ld-03.6
L10-c-ld-08.3	L11-c-ld-05.8	L12-c-ld-16.4	L13-c-ld-07.2	L14-c-ld-10.0	L15-c-ld-06.4	L16-c-ld-06.5	L17-c-ld-05.0
L10-c-ld-13.3	L11-c-ld-06.8		L13-c-ld-09.5	L14-c-ld-12.2	L15-c-ld-09.3	L16-c-ld-07.9	L17-c-ld-07.0
L10-c-ld-14.6	L11-c-ld-06.9		L13-c-ld-11.0		L15-c-ld-09.6	L16-c-ld-09.8	L17-c-ld-07.5
	L11-c-ld-09.0						L17-c-ld-08.0
							L17-c-ld-10.4
							L17-c-ld-12.8
Parole spontanée							
L10-c-s-04.2	L11-c-s-06.7	L12-c-s-07.8	L13-c-s-01.0	L14-c-s-02.8	L15-c-s-01.7	L16-c-s-03.7	L17-c-s-02.0
L10-c-s-04.7	L11-c-s-06.9	L12-c-s-10.6	L13-c-s-01.1	L14-c-s-03.8	L15-c-s-04.9	L16-c-s-04.8	L17-c-s-02.2
L10-c-s-05.0	L11-c-s-07.6	L12-c-s-10.9	L13-c-s-03.1	L14-c-s-04.8	L15-c-s-05.0	L16-c-s-04.9	L17-c-s-03.1
L10-c-s-05.2	L11-c-s-08.3	L12-c-s-11.0	L13-c-s-06.4	L14-c-s-07.5	L15-c-s-05.5	L16-c-s-05.0	L17-c-s-03.6
L10-c-s-05.7	L11-c-s-08.7	L12-c-s-11.1	L13-c-s-09.1	L14-c-s-08.5	L15-c-s-07.3	L16-c-s-05.1	L17-c-s-04.4
L10-c-s-05.8	L11-c-s-09.0	L12-c-s-11.5	L13-c-s-10.7	L14-c-s-10.1	L15-c-s-07.4	L16-c-s-05.2	L17-c-s-04.5
L10-c-s-06.1	L11-c-s-09.2	L12-c-s-11.9	L13-c-s-12.5	L14-c-s-10.9	L15-c-s-07.5	L16-c-s-05.3	L17-c-s-04.6
L10-c-s-06.2	L11-c-s-10.0	L12-c-s-12.1	L13-c-s-12.6	L14-c-s-11.4	L15-c-s-08.2	L16-c-s-07.4	L17-c-s-05.5
L10-c-s-06.3	L11-c-s-10.2	L12-c-s-12.2	L13-c-s-13.3	L14-c-s-12.6	L15-c-s-08.5	L16-c-s-07.7	L17-c-s-06.2
L10-c-s-06.5	L11-c-s-10.8	L12-c-s-12.3	L13-c-s-13.8	L14-c-s-13.7	L15-c-s-08.7	L16-c-s-07.9	L17-c-s-06.6
L10-c-s-06.6	L11-c-s-11.1	L12-c-s-13.7	L13-c-s-14.7	L14-c-s-13.9	L15-c-s-08.8	L16-c-s-08.2	L17-c-s-07.8
L10-c-s-07.1	L11-c-s-11.2	L12-c-s-14.5	L13-c-s-14.8	L14-c-s-14.0	L15-c-s-09.0	L16-c-s-08.5	L17-c-s-08.0
L10-c-s-07.6	L11-c-s-11.4	L12-c-s-14.9	L13-c-s-15.1	L14-c-s-14.1	L15-c-s-09.2		
L10-c-s-07.7	L11-c-s-13.7	L12-c-s-16.4	L13-c-s-15.6	L14-c-s-14.2	L15-c-s-09.5		
L10-c-s-07.8	L11-c-s-17.2	L12-c-s-18.1	L13-c-s-15.7	L14-c-s-17.0	L15-c-s-10.7		
L10-c-s-07.9	L11-c-s-17.7	L12-c-s-18.4	L13-c-s-15.8	L14-c-s-18.0	L15-c-s-11.1		
L10-c-s-10.1	L11-c-s-17.8	L12-c-s-19.4	L13-c-s-17.6	L14-c-s-18.3	L15-c-s-11.6		
L10-c-s-16.2			L13-c-s-17.8	L14-c-s-20.7	L15-c-s-13.2		
			L13-c-s-19.7	L14-c-s-21.5	L15-c-s-14.5		
			L13-c-s-21.1	L14-c-s-23.0			
				L14-c-s-24.2			

A.VI.3.d. Enregistrements de comparaison des locuteurs L18 - L56

Locuteur 18	Locuteur 19	Locuteur 20	Locuteur 22	Locuteur 39	Locuteur 40	Locuteur 41	Locuteur 56
Messages anonymes							
L18-c-ad-09.5	L19-c-ad-07.1	L20-c-ad-08.5	L22-c-ad-08.0	L39-c-ad-12.2	L40-c-ad13.3	L41-c-ad-06.4	L56-c-ad-10.7
L18-c-an-09.1	L19-c-an-06.6	L20-c-an-09.7	L22-c-an-07.8	L39-c-ad-30.2	L40-c-an12.2	L41-c-an-06.1	L56-c-an-08.1
				L39-c-an-10.4		L41-c-an-08.6	
Simulation de dialogues							
L18-c-d1-01.3	L19-c-d1-01.2	L20-c-d1-01.5	L22-c-d1-01.7	L39-c-d1-02.0	L40-c-d1-01.4	L41-c-d1-01.3	L56-c-d1-01.4
L18-c-d1-01.4	L19-c-d1-01.4	L20-c-d1-02.0	L22-c-d1-01.3	L39-c-d1-02.1	L40-c-d1-01.8	L41-c-d1-01.4	L56-c-d1-01.5
L18-c-d1-01.7	L19-c-d1-01.8	L20-c-d1-02.1	L22-c-d1-01.9	L39-c-d1-03.0	L40-c-d1-02.2	L41-c-d1-01.5	L56-c-d1-01.8
L18-c-d1-02.1	L19-c-d1-02.0	L20-c-d1-02.5	L22-c-d1-02.4	L39-c-d1-03.2	L40-c-d1-02.4	L41-c-d1-02.1	L56-c-d1-04.2
L18-c-d1-05.3	L19-c-d2-00.8	L20-c-d2-01.3	L22-c-d2-01.0	L39-c-d2-01.4	L40-c-d2-01.3	L41-c-d1-02.4	L56-c-d2-01.9
L18-c-d2-02.1	L19-c-d2-01.8	L20-c-d2-02.2	L22-c-d2-01.8	L39-c-d2-02.5	L40-c-d2-02.7	L41-c-d2-02.1	L56-c-d2-02.2
L18-c-d2-02.2	L19-c-d2-01.9	L20-c-d2-02.5	L22-c-d2-02.5	L39-c-d2-04.0	L40-c-d2-03.3	L41-c-d2-02.3	L56-c-d2-02.5
L18-c-d2-02.4	L19-c-d2-04.6	L20-c-d2-04.5	L22-c-d2-05.1	L39-c-d2-06.5	L40-c-d2-05.9	L41-c-d2-05.8	L56-c-d2-02.7
Lecture déguisée							
L18-c-ld-05.0	L19-c-ld-03.4	L20-c-ld-04.6	L22-c-ld-02.9	L39-c-ld-11.0	L40-c-ld05.2	L41-c-ld-03.6	L56-c-ld-03.9
L18-c-ld-06.9	L19-c-ld-04.7	L20-c-ld-06.8	L22-c-ld-04.2	L39-c-ld-11.7	L40-c-ld07.7	L41-c-ld-06.0	L56-c-ld-06.3
L18-c-ld-08.4	L19-c-ld-05.7	L20-c-ld-06.9	L22-c-ld-06.3	L39-c-ld-19.3	L40-c-ld13.4	L41-c-ld-06.1	L56-c-ld-07.1
L18-c-ld-12.3	L19-c-ld-08.8	L20-c-ld-11.1	L22-c-ld-06.4		L40-c-ld23.9	L41-c-ld-08.9	L56-c-ld-11.4
L18-c-ld-14.7	L19-c-ld-09.1	L20-c-ld-11.4	L22-c-ld-06.9			L41-c-ld-09.6	L56-c-ld-13.6
			L22-c-ld-08.1				
Parole spontanée							
L18-c-s-02.3	L19-c-s-01.0	L20-c-s-04.2	L22-c-s-02.0	L39-c-s-02.0	L40-c-s-05.3	L41-c-s-01.5	L56-c-s-05.5
L18-c-s-02.9	L19-c-s-01.5	L20-c-s-04.3	L22-c-s-02.8	L39-c-s-02.4	L40-c-s-05.9	L41-c-s-01.8	L56-c-s-09.8
L18-c-s-03.0	L19-c-s-02.8	L20-c-s-08.1	L22-c-s-02.9	L39-c-s-02.8	L40-c-s-06.0	L41-c-s-015.9	L56-c-s-10.0
L18-c-s-05.0	L19-c-s-03.2	L20-c-s-08.4	L22-c-s-03.0	L39-c-s-03.4	L40-c-s-07.5	L41-c-s-02.0	L56-c-s-12.8
L18-c-s-06.5	L19-c-s-03.7	L20-c-s-08.9	L22-c-s-03.1	L39-c-s-03.8	L40-c-s-08.4	L41-c-s-02.1	L56-c-s-12.9
L18-c-s-06.6	L19-c-s-03.8	L20-c-s-09.4	L22-c-s-03.4	L39-c-s-04.3	L40-c-s-08.6	L41-c-s-02.2	L56-c-s-13.4
L18-c-s-06.7	L19-c-s-05.0	L20-c-s-11.1	L22-c-s-03.5	L39-c-s-04.5	L40-c-s-09.6	L41-c-s-02.4	L56-c-s-13.6
L18-c-s-09.4	L19-c-s-05.5	L20-c-s-12.4	L22-c-s-03.8	L39-c-s-05.0	L40-c-s-10.0	L41-c-s-03.3	L56-c-s-14.6
L18-c-s-09.7	L19-c-s-06.5	L20-c-s-13.1	L22-c-s-04.5	L39-c-s-05.3	L40-c-s-10.8	L41-c-s-03.4	L56-c-s-14.8
L18-c-s-10.8	L19-c-s-06.9	L20-c-s-13.2	L22-c-s-04.8	L39-c-s-05.4	L40-c-s-12.2	L41-c-s-03.5	L56-c-s-15.3
L18-c-s-11.8	L19-c-s-07.5	L20-c-s-14.3	L22-c-s-04.9	L39-c-s-07.3	L40-c-s-13.7	L41-c-s-04.0	L56-c-s-15.4
L18-c-s-13.2	L19-c-s-07.8	L20-c-s-14.7	L22-c-s-05.0	L39-c-s-07.6	L40-c-s-14.0	L41-c-s-04.9	L56-c-s-15.5
L18-c-s-14.1	L19-c-s-08.7	L20-c-s-15.4	L22-c-s-05.3	L39-c-s-07.7	L40-c-s-14.1	L41-c-s-05.0	L56-c-s-16.3
L18-c-s-14.4	L19-c-s-10.3	L20-c-s-16.3	L22-c-s-06.2	L39-c-s-08.8	L40-c-s-14.4	L41-c-s-05.6	L56-c-s-17.1
L18-c-s-14.7	L19-c-s-13.0	L20-c-s-16.5	L22-c-s-06.3	L39-c-s-09.1	L40-c-s-17.4	L41-c-s-06.5	L56-c-s-17.8
L18-c-s-18.1	L19-c-s-13.4	L20-c-s-17.4	L22-c-s-06.4	L39-c-s-09.2	L40-c-s-17.6	L41-c-s-06.6	L56-c-s-18.3
L18-c-s-18.4		L20-c-s-18.3		L39-c-s-09.5	L40-c-s-19.4	L41-c-s-07.0	L56-c-s-20.0
L18-c-s-19.0		L20-c-s-18.7		L39-c-s-09.9	L40-c-s-19.5	L41-c-s-07.3	
L18-c-s-19.8				L39-c-s-10.2	L40-c-s-22.6	L41-c-s-08.7	
L18-c-s-21.4				L39-c-s-10.3	L40-c-s-27.9	L41-c-s-09.0	
L18-c-s-21.5				L39-c-s-11.1		L41-c-s-09.8	
L18-c-s-21.6				L39-c-s-11.5		L41-c-s-10.2	
L18-c-s-22.5				L39-c-s-11.6		L41-c-s-10.4	
L18-c-s-25.1				L39-c-s-11.8		L41-c-s-10.7	
L18-c-s-25.7				L39-c-s-12.1		L41-c-s-12.4	
L18-c-s-26.3				L39-c-s-12.3		L41-c-s-12.7	
				L39-c-s-12.8		L41-c-s-13.0	
				L39-c-s-12.9		L41-c-s-13.2	
				L39-c-s-13.4		L41-c-s-13.6	
				L39-c-s-13.5		L41-c-s-13.8	
				L39-c-s-13.9		L41-c-s-13.9	

Locuteur 18	Locuteur 19	Locuteur 20	Locuteur 22	Locuteur 39	Locuteur 40	Locuteur 41	Locuteur 56
Parole spontanée							
				L39-c-s-14.0		L41-c-s-14.5	
				L39-c-s-14.1		L41-c-s-15.0	
				L39-c-s-15.3		L41-c-s-15.1	
				L39-c-s-16.0		L41-c-s-15.3	
				L39-c-s-17.0		L41-c-s-16.9	
				L39-c-s-17.9		L41-c-s-17.0	
				L39-c-s-18.2		L41-c-s-18.6	
				L39-c-s-18.3		L41-c-s-18.8	
				L39-c-s-18.6		L41-c-s-18.9	
				L39-c-s-19.4		L41-c-s-20.8	
				L39-c-s-19.5		L41-c-s-21.8	
						L41-c-s-23.8	

Enregistrements de test

(-test) indique les sessions d'enregistrement de test

(-ad) indique un message anonyme avec déguisement libre

(-an) indique un message anonyme sans déguisement de la voix

A.VI.3.e. Enregistrements de test des locutrices L00 - L09

Locutrice 00	Locutrice 01	Locutrice 04	Locutrice 05	Locutrice 06	Locutrice 07	Locutrice 08	Locutrice 09
Messages anonymes							
L00-test-ad	L01-test-ad	L04-test-ad	L05-test-ad	L06-test-ad	L07-test-ad	L08-test-ad	L09-test-ad
L00-test-an	L01-test-an	L04-test-an	L05-test-an	L06-test-an	L07-test-an	L08-test-an	L09-test-an
Téléphone cellulaire							
L00-test-cellulaire	L01-test-cellulaire	L04-test-cellulaire	L05-test-cellulaire	L06-test-cellulaire	L07-test-cellulaire	L08-test-cellulaire	L09-test-cellulaire
Tests bruités							
L00-test1-0dB	L01-test1-0dB	L04-test1-0dB	L05-test1-0dB	L06-test1-0dB	L07-test1-0dB	L08-test1-0dB	L09-test1-0dB
L00-test1-3dB	L01-test1-3dB	L04-test1-3dB	L05-test1-3dB	L06-test1-3dB	L07-test1-3dB	L08-test1-3dB	L09-test1-3dB
L00-test1-6dB	L01-test1-6dB	L04-test1-6dB	L05-test1-6dB	L06-test1-6dB	L07-test1-6dB	L08-test1-6dB	L09-test1-6dB
L00-test1-9dB	L01-test1-9dB	L04-test1-9dB	L05-test1-9dB	L06-test1-9dB	L07-test1-9dB	L08-test1-9dB	L09-test1-9dB
L00-test1-12dB	L01-test1-12dB	L04-test1-12dB	L05-test1-12dB	L06-test1-12dB	L07-test1-12dB	L08-test1-12dB	L09-test1-12dB
L00-test1-18dB	L01-test1-18dB	L04-test1-18dB	L05-test1-18dB	L06-test1-18dB	L07-test1-18dB	L08-test1-18dB	L09-test1-18dB
L00-test1-24dB	L01-test1-24dB	L04-test1-24dB	L05-test1-24dB	L06-test1-24dB	L07-test1-24dB	L08-test1-24dB	L09-test1-24dB
L00-test1-30dB	L01-test1-30dB	L04-test1-30dB	L05-test1-30dB	L06-test1-30dB	L07-test1-30dB	L08-test1-30dB	L09-test1-30dB
Test analogique							
L00-test1-analogique	L01-test1-analogique	L04-test1-analogique	L05-test1-analogique	L06-test1-analogique	L07-test1-analogique	L08-test1-analogique	L09-test1-analogique
Parole spontanée							
L00-test1	L01-test1	L04-test1	L05-test1	L06-test1	L07-test1	L08-test1	L09-test1
L00-test2	L01-test2	L04-test2	L05-test2	L06-test2	L07-test2	L08-test2	L09-test2
L00-test3	L01-test3	L04-test3	L05-test3	L06-test3	L07-test3	L08-test3	L09-test3
L00-test4	L01-test4	L04-test4	L05-test4	L06-test4	L07-test4	L08-test4	L09-test4
L00-test5	L01-test5	L04-test5	L05-test5	L06-test5	L07-test5	L08-test5	L09-test5

A.VI.3.f. Enregistrements de test des locutrices L32 - L59

Locutrice 32	Locutrice 33	Locutrice 44	Locutrice 49	Locutrice 54	Locutrice 55	Locutrice 58	Locutrice 59
Messages anonymes							
L32-test-ad	L33-test-ad	L44-test-ad	L49-test-ad	L54-test-ad	L55-test-ad	L58-test-ad	L59-test-ad
L32-test-an	L33-test-an	L44-test-an	L49-test-an	L54-test-an	L55-test-an	L58-test-an	L59-test-an
Téléphone cellulaire							
L32-test-cellulaire	L33-test-cellulaire	L44-test-cellulaire	L49-test-cellulaire	L54-test-cellulaire	L55-test-cellulaire	L58-test-cellulaire	L59-test-cellulaire
Tests bruités							
L32-test1-0dB	L33-test1-0dB	L44-test1-0dB	L49-test1-0dB	L54-test1-0dB	L55-test1-0dB	L58-test1-0dB	L59-test1-0dB
L32-test1-3dB	L33-test1-3dB	L44-test1-3dB	L49-test1-3dB	L54-test1-3dB	L55-test1-3dB	L58-test1-3dB	L59-test1-3dB
L32-test1-6dB	L33-test1-6dB	L44-test1-6dB	L49-test1-6dB	L54-test1-6dB	L55-test1-6dB	L58-test1-6dB	L59-test1-6dB
L32-test1-9dB	L33-test1-9dB	L44-test1-9dB	L49-test1-9dB	L54-test1-9dB	L55-test1-9dB	L58-test1-9dB	L59-test1-9dB
L32-test1-12dB	L33-test1-12dB	L44-test1-12dB	L49-test1-12dB	L54-test1-12dB	L55-test1-12dB	L58-test1-12dB	L59-test1-12dB
L32-test1-18dB	L33-test1-18dB	L44-test1-18dB	L49-test1-18dB	L54-test1-18dB	L55-test1-18dB	L58-test1-18dB	L59-test1-18dB
L32-test1-24dB	L33-test1-24dB	L44-test1-24dB	L49-test1-24dB	L54-test1-24dB	L55-test1-24dB	L58-test1-24dB	L59-test1-24dB
L32-test1-30dB	L33-test1-30dB	L44-test1-30dB	L49-test1-30dB	L54-test1-30dB	L55-test1-30dB	L58-test1-30dB	L59-test1-30dB
Test analogique							
L32-test1-analogique	L33-test1-analogique	L44-test1-analogique	L49-test1-analogique	L54-test1-analogique	L55-test1-analogique	L58-test1-analogique	L59-test1-analogique
Parole spontanée							
L32-test1	L33-test1	L44-test1	L49-test1	L54-test1	L55-test1	L58-test1	L59-test1
L32-test2	L33-test2	L44-test2	L49-test2	L54-test2	L55-test2	L58-test2	L59-test2
L32-test3	L33-test3	L44-test3	L49-test3	L54-test3	L55-test3	L58-test3	L59-test3
L32-test4	L33-test4	L44-test4	L49-test4	L54-test4	L55-test4	L58-test4	L59-test4
L32-test5	L33-test5	L44-test5	L49-test5	L54-test5	L55-test5	L58-test5	L59-test5

74A.VI.3.g. Enregistrements de test des locuteurs L10 - L17

Locuteur 10	Locuteur 11	Locuteur 12	Locuteur 13	Locuteur 14	Locuteur 15	Locuteur 16	Locuteur 17
Messages anonymes							
L10-test-ad	L11-test-ad	L12-test-ad	L13-test-ad	L14-test-ad	L15-test-ad	L16-test-ad	L17-test-ad
L10-test-an	L11-test-an	L12-test-an	L13-test-an	L14-test-an	L15-test-an	L16-test-an	L17-test-an
Téléphone cellulaire							
L10.-test-cellulaire	L11.-test-cellulaire	L12.-test-cellulaire	L13.-test-cellulaire	L14.-test-cellulaire	L15.-test-cellulaire	L16.-test-cellulaire	L17.-test-cellulaire
Tests bruités							
L10.-test1-0dB	L11.-test1-0dB	L12.-test1-0dB	L13.-test1-0dB	L14.-test1-0dB	L15.-test1-0dB	L16.-test1-0dB	L17.-test1-0dB
L10.-test1-3dB	L11.-test1-3dB	L12.-test1-3dB	L13.-test1-3dB	L14.-test1-3dB	L15.-test1-3dB	L16.-test1-3dB	L17.-test1-3dB
L10.-test1-6dB	L11.-test1-6dB	L12.-test1-6dB	L13.-test1-6dB	L14.-test1-6dB	L15.-test1-6dB	L16.-test1-6dB	L17.-test1-6dB
L10.-test1-9dB	L11.-test1-9dB	L12.-test1-9dB	L13.-test1-9dB	L14.-test1-9dB	L15.-test1-9dB	L16.-test1-9dB	L17.-test1-9dB
L10.-test1-12dB	L11.-test1-12dB	L12.-test1-12dB	L13.-test1-12dB	L14.-test1-12dB	L15.-test1-12dB	L16.-test1-12dB	L17.-test1-12dB
L10.-test1-18dB	L11.-test1-18dB	L12.-test1-18dB	L13.-test1-18dB	L14.-test1-18dB	L15.-test1-18dB	L16.-test1-18dB	L17.-test1-18dB
L10.-test1-24dB	L11.-test1-24dB	L12.-test1-24dB	L13.-test1-24dB	L14.-test1-24dB	L15.-test1-24dB	L16.-test1-24dB	L17.-test1-24dB
L10.-test1-30dB	L11.-test1-30dB	L12.-test1-30dB	L13.-test1-30dB	L14.-test1-30dB	L15.-test1-30dB	L16.-test1-30dB	L17.-test1-30dB
Test analogique							
L10-test1-analogique	L11-test1-analogique	L12-test1-analogique	L13-test1-analogique	L14-test1-analogique	L15-test1-analogique	L16-test1-analogique	L17-test1-analogique
Parole spontanée							
L10-test1	L11-test1	L12-test1	L13-test1	L14-test1	L15-test1	L16-test1	L17-test1
L10-test2	L11-test2	L12-test2	L13-test2	L14-test2	L15-test2	L16-test2	L17-test2
L10-test3	L11-test3	L12-test3	L13-test3	L14-test3	L15-test3	L16-test3	L17-test3
L10-test4	L11-test4	L12-test4	L13-test4	L14-test4	L15-test4	L16-test4	L17-test4
L10-test5	L11-test5	L12-test5	L13-test5	L14-test5	L15-test5	L16-test5	L17-test5

A.VI.3.h. Enregistrements de test des locuteurs L18 - L56

Locuteur 18	Locuteur 19	Locuteur 20	Locuteur 22	Locuteur 39	Locuteur 40	Locuteur 41	Locuteur 56
Messages anonymes							
L18-test-ad	L19-test-ad	L20-test-ad	L22-test-ad	L39-test-ad	L40-test-ad	L41-test-ad	L56-test-ad
L18-test-ad1	L19-test-ad1	L20-test-ad1	L22-test-ad1	L39-test-an	L40-test-an	L41-test-ad1	L56-test-an
L18-test-ad2	L19-test-ad2	L20-test-ad2	L22-test-ad2			L41-test-ad2	
L18-test-an	L19-test-ad3	L20-test-an	L22-test-an			L41-test-an	
L18-test-an1	L19-test-an		L22-test-an1			L41-test-an1	
L18-test-an2	L19-test-an1		L22-test-an2			L41-test-an2	
L18-test-an3	L19-test-an2					L41-test-an3	
L18-test-an4	L19-test-an3						
L18-test-an5							
Téléphone cellulaire							
L18-test-cellulaire	L19-test-cellulaire	L20-test-cellulaire	L22-test-cellulaire	L39-test-cellulaire	L40-test-cellulaire	L41-test-cellulaire	L56-test-cellulaire
Tests bruités							
L18-test1-0dB	L19-test1-0dB	L20-test1-0dB	L22-test1-0dB	L39-test1-0dB	L40-test1-0dB	L41-test1-0dB	L56-test1-0dB
L18-test1-3dB	L19-test1-3dB	L20-test1-3dB	L22-test1-3dB	L39-test1-3dB	L40-test1-3dB	L41-test1-3dB	L56-test1-3dB
L18-test1-6dB	L19-test1-6dB	L20-test1-6dB	L22-test1-6dB	L39-test1-6dB	L40-test1-6dB	L41-test1-6dB	L56-test1-6dB
L18-test1-9dB	L19-test1-9dB	L20-test1-9dB	L22-test1-9dB	L39-test1-9dB	L40-test1-9dB	L41-test1-9dB	L56-test1-9dB
L18-test1-12dB	L19-test1-12dB	L20-test1-12dB	L22-test1-12dB	L39-test1-12dB	L40-test1-12dB	L41-test1-12dB	L56-test1-12dB
L18-test1-18dB	L19-test1-18dB	L20-test1-18dB	L22-test1-18dB	L39-test1-18dB	L40-test1-18dB	L41-test1-18dB	L56-test1-18dB
L18-test1-24dB	L19-test1-24dB	L20-test1-24dB	L22-test1-24dB	L39-test1-24dB	L40-test1-24dB	L41-test1-24dB	L56-test1-24dB
L18-test1-30dB	L19-test1-30dB	L20-test1-30dB	L22-test1-30dB	L39-test1-30dB	L40-test1-30dB	L41-test1-30dB	L56-test1-30dB
Test analogique							
L18-test1-analogique	L19-test1-analogique	L20-test1-analogique	L22-test1-analogique	L39-test1-analogique	L40-test1-analogique	L41-test1-analogique	L56-test1-analogique
Parole spontanée							
L18-test1	L19-test1	L20-test1	L22-test1	L39-test1	L40-test1	L41-test1	L56-test1
L18-test2	L19-test2	L20-test2	L22-test2	L39-test2	L40-test2	L41-test2	L56-test2
L18-test3	L19-test3	L20-test3	L22-test3	L39-test3	L40-test3	L41-test3	L56-test3
L18-test4	L19-test4	L20-test4	L22-test4	L39-test4	L40-test4	L41-test4	L56-test4
L18-test5	L19-test5	L20-test5	L22-test5	L39-test5	L40-test5	L41-test5	L56-test5

BIBLIOGRAPHIE

BIBLIOGRAPHIE

- ABBERTON, E. ; FOURCIN, A. J. ; (1978) ; « *Intonation and speaker identification* » ; Language and Speech ; vol. 21 ; pp. 305 - 318.
- AIGRIN, P. ; (1996) ; « *Experts-alibis, experts piégés, experts responsables* » ; Le Monde ; p. 15 ; 23 octobre.
- AITKEN, C. G. G. ; (1995) ; « *Statistics and the evaluation of evidence for forensic scientists* » ; John Wiley & Sons, Chichester.
- ALEXANDERSON, R. ; (1997) ; « *Communication personnelle* » ; 30 janvier.
- ALLPORT, G. W. ; (1963) ; « *Pattern and growth in personality* » ; Holt, Rinehart & Winston, New York.
- ALLPORT, G. W. ; CANTRIL, H. ; (1934) ; « *Judging personality from voice* » ; J. Soc. Psychol. ; vol. 5 ; pp. 37 - 55.
- ANGHELESCU, I. ; (1974) ; « *Méthode d'identification des personnes d'après la voix et la manière de parler en roumain* » ; RIPC ; vol. 28 ; no. 274 ; pp. 2 - 8.
- ANGHELESCU, I. ; (1985) ; « *L'expertise criminalistique de la voix* » ; RIPC ; vol. 40 ; no. 390 ; pp. 180 - 185.
- ANONYME ; (1991) ; « *A la recherche d'une signature vocale* » ; Pol. Nat. Fr. ; no. 6 ; pp. 14 - 16.
- ARGYLE, M. ; (1976) ; « *The psychology of interpersonal behaviour* » ; 2^e éd ; Penguin Books, Harmondsworth.
- ARISTOTE ; (384 - 322 av. J.-C.) ; « *De Interpretatione* ».
- ASSALEH, K. T. ; MAMMONE, R. J. ; (1994) ; « *Robust cepstral features for speaker identification* » ; Proceedings - ICASSP ; vol. 1 ; pp. I-129 - 132.
- ATAL, B. S. ; (1968) ; « *Automatic speaker recognition based on speech contours* » ; Ph. D. Thesis, Polytech. Int. Brooklyn, NY.
- ATAL, B. S. ; (1974) ; « *Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification* » ; J. Acoust. Soc. Am. ; vol. 55 ; pp. 1304 - 1312.
- ATAL, B. S. ; (1976) ; « *Automatic recognition of speakers from their voices* » ; Proc. IEEE ; vol. 64 ; no. 4 ; p. 460.
- ATKINSON, J. E. ; (1976) ; « *Inter and intra-speaker variability in fundamental voice frequency* » ; J. Acoustic. Soc. Am. ; vol. 60 ; no. 2 ; pp. 440 - 455.
- ATWOOD, W. ; HOLLIEN, H. ; (1986) ; « *Stress monitoring by polygraph for research purposes* » ; Polygraph ; vol. 15 ; pp. 47 - 56.
- BALDWIN, J. ; FRENCH, P. ; (1990) ; « *Forensic Phonetics* » ; Pinter, London.
- BASZTURA, C. ; JURKIEWICZ, J. ; (1978) ; « *The zero-crossing analysis of a speech signal in the short-term method of automatic speaker recognition* » ; Archives of acoustics ; vol. 3 ; no. 3 ; pp. 185 - 196.
- BASZTURA, C. ; MAJEWSKI, W. ; (1978) ; « *The application of long-term analysis of the zero-crossing analysis of a speech signal in automatic speaker identification* » ; Archives of acoustics ; ; vol. 1 ; no. 3 pp. 3 - 15.

- BECKER, R. W. ; CLARKE, F. R. ; POZA, F. T. ; YOUNG, J. R. ; (1973) ; « *A semi-automatic speaker recognition system* » ; U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice, Washington.
- BERNASCONI, C. ; (1990) ; « *On instantaneous and transitional spectral information for text-dependent speaker verification* » ; Speech Communication ; vol. 9 ; no. 2 ; pp. 129 - 139.
- BERTILLON, A. ; (1881) ; « *Une application de l'anthropométrie sur un procédé d'identification* » ; Annales de Démographie Internationale ; G. Masson, Paris.
- BERTILLON, A. ; (1893) ; « *Renseignements descriptifs* » IN: « *Identification anthropométrique - instructions signalétiques* » ; Imprimerie administrative, Melun ; pp. 103 - 105.
- BIMBOT, F. ; (1993) ; « *Assessment methodology for speaker identification and verification systems* » ; SAM-A Esprit Project 6819, report I-9, task 2500.
- BIMBOT, F. ; CHOLLET, G. ; PAOLONI, A. ; (1994) ; « *Assessment methodology for speaker identification and verification systems: an overview of SAM-A Esprit Project 6819 - Task 2500* » ; Proceedings of ESCA Workshop on automatic speaker recognition, identification and verification ; pp. 75 - 82.
- BIMBOT, F. ; MAGRIN-CHAGNOLLEAU, I. ; MATHAN, L. ; (1995) ; « *Second-order statistical measures for text-independent speaker identification* » ; Speech Communication ; pp. 177 -192.
- BIMBOT, F. ; MATHAN, L. ; (1994) ; « *Second-order statistical measures for text-independent speaker identification* » ; Proceedings of ESCA Workshop on automatic speaker recognition, identification and verification ; pp. 51 - 54.
- BLACK, B. ; AYALA, F. J. ; SAFFRAN-BRINKS, C. ; (1994) ; « *Science and the law in the wake of Daubert: A new search for scientific knowledge* » ; Texas Law Review ; vol. 72 ; no. 4 ; pp. 715 - 802.
- BLACK, J. W. ; LASHBROOK, W. B. ; NASH, E. W. ; OYER, H. J. ; PEDREY, C. ; TOSI, O. I. ; TRUBY, H. ; (1974) ; « *Reply to: Speaker identification by speech spectrograms: some further observations* » ; J. Acoust. Soc. Am. ; vol. 54 ; pp. 535 - 537.
- BLOCK, E. B. ; (1975) ; « *Voiceprinting : how the law can read the voice of crime* » ; D. McKay Co., New York.
- BOË, L. J. ; (1998) ; « *L'identification juridique de la voix : le cas français - Historique, problématique et propositions* » ; Proceedings of RLA2C Workshop : « *Speaker Recognition and its Commercial and Forensic Applications* » ; pp. 222 - 239.
- BÖHME, G. ; HECKER, G. ; (1970) ; « *Gerontologische Untersuchungen über Stimmumfang und Sprechstimmlage* » ; Folia Phoniatria ; vol. 22 ; pp. 176 - 184.
- BOITE, R. ; KUNT, M. ; (1987) ; « *Traitement de la parole* » ; Presses polytechniques romandes, Lausanne.
- BOLT, R. H. ; COOPER, F. S. ; DAVID, E. E. ; DENES, P. B. ; PICKETT, J. M. ; STEVENS, K. N. ; (1969) ; « *Speaker identification by speech spectrograms* » ; Science ; pp. 338 - 343.
- BOLT, R. H. ; COOPER, F. S. ; DAVID, E. E. ; DENES, P. B. ; PICKETT, J. M. ; STEVENS, K. N. ; (1970) ; « *Speaker identification by speech spectrograms: A scientists' view of its reliability for legal purposes* » ; J. Acoustic. Soc. Am. ; vol. 47 ; no. 2 ; pp. 597 - 612.

- BOLT, R. H. ; COOPER, F. S. ; DAVID, E. E. ; DENES, P. B. ; PICKETT, J. M. ; STEVENS, K. N. ;** (1973) ; « *Speaker identification by speech spectrograms: some further observations* » ; J. Acoust. Soc. Am. ; vol. 54 ; pp. 531 - 534.
- BOLT, R. H. ; COOPER, F. S. ; GREEN, D. ; HAMLET, S. L. ; HOGAN, D. L. ; MC KNIGHT, J. G. ; PIKETT, J. M. ; TOSI, O. ; UNDERWOOD, B. D. ;** (1979) ; « *On the theory and practice of voice identification* » ; National Academy of Sciences, Washington.
- BONAVENTURA, M. ;** (1935) ; « *Ausdruck der Persönlichkeit in der Sprechstimme und im Phonogramm* » ; Arch. Ges. Psychol. ; vol. 94 ; pp. 501 - 570.
- BORDERS, W. ;** (1966) ; « *Voiceprint allowed as evidence: Ruling called first of this kind* » ; The New York Times, April 12th.
- BOVES, L. ;** (1998) ; « *Commercial applications of speaker verification: overview and critical success factors* » ; Proceedings of RLA2C Workshop: « *Speaker Recognition and its Commercial and Forensic Applications* » ; pp. 150 - 160.
- BRADSHAW, J. ; NETTLETON, N. ;** (1983) ; « *Human Cerebral Asymmetry* » ; Prentice-Hall, Englewood Cliffs, NJ.
- BRAUN, A. ;** (1994) ; « *The effect of cigarette smoking on vocal parameters* » ; Proceedings of ESCA Workshop on automatic speaker recognition, identification and verification ; pp. 161- 164.
- BRAUN, A. ;** (1995) ; « *Procedures and Perspectives in Forensic Phonetics* » ; Proceedings of the XIIth International Congress of Phonetic Sciences, Stockholm ; vol. 3 ; pp. 146 - 153.
- BRAUN, A. ;** (1996) ; « *Age estimation by different listener groups* » ; Forensic Linguistics ; vol. 3 ; no. 1 ; pp. 50 - 64.
- BRAUN, A. ;** (1998) ; « *Voice Analysis* » ; rapport présenté lors de la 12^{ème} Conférence Triennale d'Interpol sur les sciences forensiques à Lyon.
- BRAUN, A. ; RIETVELD, T. ;** (1995) ; « *The Influence of smoking habits on perceived age* » ; Proceedings of the XIIth International Congress of Phonetic Sciences, Stockholm ; vol. 1 ; pp. 294 - 297.
- BRICKER, P. ; PRUZANSKY, S. ;** (1966) ; « *Effects of stimulus content and duration on talker identification* » ; J. Acoustic. Soc. Am. ; vol. 40 ; pp. 1441 - 1449.
- BRICKER, P. ; GNANADESIKAN, R. ; MATHEWS, M. W. ; PRUZANSKI, S. ; TUKEY, P. A. ; WATCHER, K. W. ; WARNER, J. L. ;** (1971) ; « *Statistical techniques for talker identification* » ; Bell Sys. Tech. Journal ; vol. 50 ; pp. 1427 - 1454.
- BRICKER, P. ; PRUZANSKY, S. ;** (1976) ; « *Speaker recognition* » ; IN : « *Contemporary Issues in Experimental Phonetics* » (ed.: Lass, N. J.) ; New York: Academic Press ; pp. 295 -326.
- BROEDERS, A. P. A. ;** (1995) ; « *The role of automatic speaker recognition techniques in forensic investigations* » ; Proceedings of the XIIth International Congress of Phonetic Sciences, Stockholm ; vol. 3 ; pp. 154 - 161.
- BROEDERS, A. P. A. ;** (1996) ; « *Earwitness identification: common grounds, disputed territory and uncharted areas* » ; Forensic Linguistics ; vol. 3 ; no. 1 ; pp. 3 - 13.
- BROWN, B. L. ;** (1974) ; « *An experimental study of the relative importance of acoustic parameters for auditory speaker recognition* » ; Language and Speech ; vol. 24 ; pp. 295 - 310.

- BROWN, R.** ; (1981) ; « *An experimental study of the relative importance of acoustic parameters for auditory speaker recognition* » ; *Language and Speech* ; vol. 24 ; pp. 295 - 310.
- BROWN, B. L.** ; **STRONG, W. J.** ; **RENCHER, A. C.** ; (1974) ; « *Fifty-four voices from two: The effects of simultaneous manipulations of rate mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech* » ; *J. Acoustic. Soc. Am.* ; vol. 55 ; pp. 313 - 318.
- BROWN, P.** ; **LEVINSON, S.** ; (1979) ; « *Social structure, groups and interactions* » IN: « *Social markers in speech* » (eds.: Scherer & Giles) ; Cambridge University Press, Cambridge.
- BRYDEN, M. P.** ; (1982) ; « *Laterality: Functional Asymmetry in the Intact Brain* » ; Academic Press, New York.
- BUNGE, E.** ; (1977) ; « *Speaker recognition by computer* » ; *Philips Technical Review* ; vol. 37 ; no. 8 ; pp. 207 - 219.
- BUNGE, E.** ; (1979) ; « *Identification judiciaire de la voix par ordinateur* » ; *Revue Int. Pol. Crim.* ; no. 332 ; pp. 254 - 270.
- BUNGE, E.** ; (1991) ; « *The role of pattern recognition in forensic science: an introduction to methods* » IN: « *Police research in the Federal Republic of Germany. 15 years research within the Bundeskriminalamt* » (eds.: Kube, E. ; Störzer, H. U. ; Clarke, R. V.) ; Springer-Verlag, Berlin ; pp. 254 - 265.
- BURKE, J. P.** ; **COLEMAN, R. O.** ; (1973) ; « *Speaker identification by naive observers using visual comparison of contour spectrograms* » ; *The Criminologist* ; vol. 8 ; no. 30 ; pp. 46 - 52.
- CALINSKI, T.** ; **KACZMAREK, Z.** ; (1968) ; « *Application of bivariate analysis of variance to some problem in phonetic research* » IN: « *Speech analysis and synthesis* » (ed.: Jassem, W.) ; Polish Academy of Sciences, Warsaw ; vol. 1 ; pp. 43 - 52.
- CALINSKI, T.** ; **JASSEM, W.** ; **KACZMAREK, Z.** ; (1970) ; « *Investigation of vowel formant frequencies as personal voice characteristics by means of multivariate analysis of variance* » IN: « *Speech analysis and synthesis* » (ed.: Jassem, W.) ; Polish Academy of Sciences, Warsaw ; vol. 2 ; pp. 7 - 39.
- CAPPE, O.** ; (1995) ; « *Etat actuel de la recherche en reconnaissance du locuteur et des applications en criminalistique* » ; rapport interne ; Ecole Nationale des Télécommunications, Département Signal, Paris.
- CATFORD, J. C.** ; (1977) ; « *Fundamental problems in phonetics* » ; Edinburgh University Press, Edinburgh.
- CHAMPOD, C.** ; (1996) ; « *Reconnaissance automatique et analyse statistique des minuties des empreintes digitales* » ; thèse de doctorat, Institut de police scientifique et de criminologie, Université de Lausanne.
- CHAMPOD, C.** ; **MEUWLY, D.** ; (1998) ; « *The inference of identity in forensic speaker recognition* » ; Proceedings of RLA2C Workshop: « *Speaker Recognition and its Commercial and Forensic Applications* » ; pp. 125 - 135 et (2000) ; *Speech Communications* ; vol. 31 ; no. 2 - 3 ; pp. 193 - 203.
- CHAMPOD, C.** ; **TARONI, F.** ; (1994) ; « *Probabilités au procès pénal: risques et solutions* » ; *Revue pénale suisse* ; vol. 112 ; no. 2 ; pp. 194 - 219.
- CHEN, M. S.** ; **LIN, P. H.** ; **WANG, H. C.** ; (1993) ; « *Speaker identification based on a matrix quantization method* » ; *IEEE Trans. ASSP* ; vol. 41 ; no. 1 ; pp. 398 - 403.
- CHEUNG, R. S.** ; **EISENSTEIN, B. A.** ; (1978) ; « *Feature selection via programming for text-independent speaker identification* » ; *Proc. IEEE ASSP* ; pp. 397 - 403.

- CLARKE, F. R. ; BECKER, R. W. ; NIXON, J. C. ; (1966) ; « *Characteristics that determine speaker recognition* » ; ESD-TR-66-636, Bedford, Mass.: Electronic systems division, Air force Systems commands, U. S. Air Force.
- CLIFFORD, B. R. ; (1980) ; « *Voice identification by human listeners: on earwitness reliability* » ; Law and Human Behaviour ; vol. 4 ; no. 4 ; pp. 373 - 394.
- CLIFFORD, B. R. ; BULL, R. H. ; RATHBORN, H. A. ; (1981) ; « *Voice identification* » ; Res. Bull. ; no. 11 ; pp. 18 - 20.
- COLEMAN, R. O. ; (1973) ; « *Speaker identification in the absence of intersubject differences in glottal source characteristics* » ; J. Acoustic. Soc. Am. ; vol. 53 ; pp. 1741 - 1743.
- COLEMAN, R. O. ; (1976) ; « *A comparison of the two vocal characteristics to the perception of maleness and femaleness in the voice* » ; J. Speech Hearing Res. ; vol. 19 ; pp. 168 - 180.
- COMPTON, A. J. ; (1963) ; « *Effects of filtering and vocal duration upon the identification of speakers aurally* » ; J. Acoustic. Soc. Am. ; vol. 35 ; pp. 1748 - 1752.
- CORSI, P. ; (1982) ; « *Speaker recognition: A survey* » IN: « *Automatic speech analysis and recognition* » (ed.: Haton J.-P.) ; D. Reidel, Dordrecht, Holland ; pp. 277 - 308.
- CURRAN, J. M. ; TRIGGS, C. M. ; BUCKLETON, J. S. ; WALSH, K. A. J. ; HICKS, T. N. ; (1998) ; « *Assessing transfer probabilities in a Bayesian interpretation of forensic glass evidence* » ; Science & Justice ; vol. 38 ; no. 1 ; pp. 15 - 21.
- CURRAN, J. ; HICKS T. N. ; BUCKLETON, J. ; (2000) ; « *Evidentiary value of glass* » ; CTC Press, Pleasanton, CA.
- CUTLER, P. E. ; THIPGEN, C. R. ; YOUNG, T. R. ; MUELLER, E. B. ; (1972) ; « *The evidentiary value of spectrographic voice identification* » ; The Journal of criminal law, criminology and police science ; vol. 63 ; no. 3 ; pp. 343 - 355.
- DAOUST, F. ; (1995) ; « *La graphologie comme moyen d'expertise judiciaire* » ; Mémoire de diplôme postgrade en expertise en documents, Institut de police scientifique et de criminologie, Université de Lausanne.
- DAS, S. K. ; MOHN, W. S. ; (1971) ; « *A scheme for speech processing in automatic speaker verification* » ; IEEE Trans. Audio Electroacoust. ; vol. AU-19 ; pp. 32 - 43.
- DAUMER, W. R. ; (1982) ; « *Subjective evaluation of several efficient speech coders* » ; IEEE Trans. Commun. ; no. April ; pp. 655 - 662.
- DAVIS, S. B. ; (1976) ; « *Computer evaluation of laryngeal pathology based on inverse filtering of speech* » ; Ph. D. Thesis, University of California, Santa Barbara.
- DE COULON, F. ; (1990) ; « *Théorie et traitement des signaux* » ; Presses polytechniques romandes, Lausanne.
- DE FINETTI, B. ; (1975) ; « *Theory of Probability* » ; Wiley & Sons ; London.
- DE MARIA, R. ; (1994) ; « *A criminals playing field* » ; Cellular Business ; vol. 11 ; no. 9 ; pp. 24.
- DE VETH, J. ; BOURLARD, H. ; (1995) ; « *Comparison of hidden Markov model techniques for automatic speaker verification in real-world conditions* » ; Speech Communication ; vol. 17 ; no. 1 - 2 ; pp. 81 -90.
- DELATTRE, P. ; (1965) ; « *Comparing the phonetic features of English, French, German and Spanish: an interim report* » ; J. Groos, Heidelberg.
- DEVIJVER, P. A. ; KITTLER, J. ; (1982) ; « *Pattern recognition: a statistical approach* » ; Prentice-Hall inc., London.

- DODDINGTON, G. R. ; (1970) ; « *A method of speaker recognition* » ; Ph. D. Thesis, University of Wisconsin, Madison.
- DODDINGTON, G. R. ; (1976) ; « *Personal identity verification using voice* » ; Proc. Electro ; pp. 22 - 24.
- DODDINGTON, G. R. ; (1979) ; « *Personal identity verification using voice* » IN: « *Automatic Speech and Speaker Recognition* » (eds: Dixon, N. R. & Martin, T. B.) ; John Wiley & Sons, New York ; pp. 385 - 397.
- DODDINGTON, G. R. ; (1985) ; « *Speaker recognition - Identify people by their voices* » ; Proc. IEEE ; vol. 73 ; no. 11 ; p. 1651.
- DODDINGTON, G. R. ; LIGGETT, W. ; MARTIN, A. ; PRZYBOCKI, M. ; REYNOLDS, D. ; (1998) ; « *Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation* » ; ICSLP ; pp. 608 - 611.
- DOHERTY, E. T. ; (1976) ; « *An evaluation of selected acoustic parameters for use in speaker identification* » ; J. phonetics ; no. 4 ; pp. 321 - 326.
- DRYGAJLO, A. ; (1999) ; « *Cours de traitement de la parole, parties I et II* » ; Département d'Électricité, École Polytechnique Fédérale de Lausanne.
- DUDA, R. O. ; HART, P. E. ; (1973) ; « *Pattern classification and scene analysis* » ; John Wiley & Sons, New York.
- EL MALIKI, M. ; DRYGAJLO, A. ; (1998) ; « *Statistical modeling and missing feature compensation for noisy speech in forensic speaker recognition* » ; Proceedings of the 8th COST 250 workshop, Ankara: « *Speaker identification by man and by machine : Directions for forensic applications* » ; pp. 39 - 45.
- ENDRESS, W. ; BAMBACH, W. ; FLOSSER, G. ; (1971) ; « *Voice spectrograms as a function of age, voice disguise and voice imitation* » ; J. Acoust. Soc. Am. ; vol. 49 ; pp. 1842 - 1848.
- ENGEL, E. ; VENETOULIAS, A. ; (1991) ; « *Monty Hall's Probability Puzzle* » ; Chance 4 ; no. 2 ; pp. 6 - 9.
- EUSTACHE, F. ; (1995) ; « *Identification et discrimination auditive: données neuropsychologiques* » ; IN: « *Perceptions et agnosies* » (eds.: Lechevalier, B. ; Eustache, F. ; Viader, F.) ; Université De Boeck , Bruxelles ; pp. 243 - 271.
- EVETT, I. W. ; (1983) ; « *What is the Probability that This Blood Came from That Person? A Meaningful Question* » ; Journal of the Forensic Science Society ; vol. 23 ; pp. 35 - 39.
- EVETT, I. W. ; (1987) ; « *On Meaningful Questions : A two Trace Transfer Problem* » ; Journal of the Forensic Science Society ; vol. 27 ; pp. 375 - 381.
- EVETT, I. W. ; (1990) ; « *The theory of interpreting scientific transfer evidence* » ; Forensic Science Progress ; vol. 4 ; pp. 141 - 179.
- EVETT, I. W. ; (1992) ; « *Interpreting of Evidence* » ; conférence présentée lors de la 10^{ème} Conférence Triennale d'Interpol sur les sciences forensiques à Lyon.
- EVETT, I. W. ; (2000) ; « *Communication personnelle* », 7 mars.
- EVETT, I. W. ; (1995) ; « *Avoiding the transposed conditional* » ; Science & Justice ; 35 ; 2 ; pp. 127 - 131.
- EVETT, I. W. ; BUCKLETON, J. S. ; (1996) ; « *Statistical analysis of STR data* » IN: « *Advances in Forensic Haemogenetics* » (eds: Carraredo, A. ; Brinkmann, B. ; Bär, W.) ; Springer-Verlag, Heidelberg ; vol. 6 ; pp. 79 - 86.

- FÄHRMANN, R. ; (1966A) ; « *Grundprobleme der Sprecherstimmverstellung und Sprechstimmvergleichung (1.Teil)* » ; Archiv für Kriminologie ; vol. 137 ; no. 1 ; pp. 25 - 32.
- FÄHRMANN, R. ; (1966B) ; « *Grundprobleme der Sprecherstimmverstellung und Sprechstimmvergleichung (2.Teil)* » ; Archiv für Kriminologie ; vol. 137 ; no. 3 ; pp. 91 - 102.
- FALCONE, M. ; DE SARIO, N. ; (1994) ; « *A PC speaker identification system for forensic use: IDEM* » ; Proceedings of ESCA Workshop on automatic speaker recognition, identification and verification ; pp. 169 - 172.
- FALCONE, M. ; PAOLONI, A. ; DE SARIO, N. ; (1995) ; « *IDEM: A software tool to study vowel formants in speaker identification* » ; Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm ; pp. 294 - 297.
- FANT, G. ; (1960) ; « *Acoustic theory of speech production* » ; MIT Press, Cambridge, MA.
- FANT, G. ; (1973) ; « *Speech, sounds and features* » ; MIT Press, Cambridge, MA.
- FAY, P. ; MIDDLETON, W. C. ; (1940) ; « *Judgement of Kretschmerian body types from the voice as transmission over a public address system* » ; J. soc. Psychol. ; pp. 151 - 162.
- FINKELSTEIN, M. O. ; FAIRLEY, W. B. ; (1970) ; « *A Bayesian Approach to Identification Evidence* » ; Harvard Law Review ; vol. 83 ; no. 3 ; pp. 489 - 517.
- FISHER, W. M. ; DODDINGTON, G. R. ; GOUDIE-MARSHALL, K. M. ; (1986) ; « *The DARPA Speech Recognition Research Database: Specification and Status* » ; Proc. DARPA Workshop on Speech Recognition, Palo Alto (CA) ; p. 93.
- FLOCH, J. L. ; MONTACIE, C. ; CARATY, M. J. ; (1994) ; « *Investigation on speaker characterization on Orphée system technics* » ; ICASSP ; p. I-149.
- FREEH, L. ; (1996) ; « *Impact of Encryption on Law Enforcement and Public Safety* » ; <http://www.fbi.gov/encrypt.htm>.
- FRENCH, P. ; (1994) ; « *An overview of forensic phonetics with particular reference to speaker identification* » ; Forensic Linguistics ; vol. 1 ; no. 2 ; pp. 169 - 181.
- FURUI, S. ; (1981A) ; « *Cepstral analysis technique for automatic speaker verification* » ; IEEE Trans. ASSP ; vol. ASSP 29 ; pp. 254 - 272.
- FURUI, S. ; (1981B) ; « *Comparison of speaker recognition methods using statistical features and dynamic features* » ; IEEE Trans. ASSP ; vol. ASSP-29 ; pp. 342 - 350.
- FURUI, S. ; (1989) ; « *Digital speech processing, synthesis, and recognition* » ; Dekker, New York.
- FURUI, S. ; (1994) ; « *An overview of speaker recognition technology* » ; Proceedings of ESCA Workshop on automatic speaker recognition, identification and verification ; pp. 1 - 9.
- FURUI, S. ; (1997) ; « *Recent advances in speaker recognition* » IN: « *Audio- and Video-Based Biometric Person Authentication* » (eds. Bigün, J. ; Chollet, G. ; Borgefors, G.) ; Springer Verlag ; Berlin ; pp. 237 - 252.
- FURUI, S. ; ITAKURA, F. ; (1973) ; « *Talker recognition by statistical features of speech sounds* » ; Electronics and Communications in Japan ; vol. 56-A ; pp. 62 - 71.

- FURUI, S. ; ITAKURA, F. ; SAITO, S. ; (1972) ; « *Talker recognition by long time averaged speech spectrum* » ; Electronics and communications in Japan ; vol. 55-A ; pp. 54 - 61.
- GALLUSSER, A. ; (1998) ; « *L'indice matériel comme moyen de preuve - sa valeur et son utilisation par les magistrats* » ; thèse de doctorat, Institut de police scientifique et de criminologie, Université de Lausanne.
- GARVIN, P. ; LADEFOGED, P. ; (1963) ; « *Speaker identification and message identification in speech recognition* » ; *Phonetica* ; vol. 9 ; no. 4 ; pp. 193 - 199.
- GAUTHIER, J. ; (1984) ; « *Enregistrement clandestin d'une conversation téléphonique et preuve pénale* », IN: « *Gedächtnisschrift für Peter Noll* » (hrsg. von Hauser, R. ; Rehberg, J. ; Stratenwerth, G.) ; pp. 333 - 340.
- GFCP: BUREAU DU GROUPE « COMMUNICATION PARLEE" DE LA SOCIETE FRANÇAISE D'ACOUSTIQUE ; (1991) ; « *About the ethics of speaker identification* » ; XXth Congrès International de Phonétique, Aix-en-Provence ; vol. 1 ; pp. 397.
- GIANELLI, P. C. ; IMWINKELRIED, E. J. ; (1986) ; « *Voice identification* » IN: « *Scientific evidence* » ; The Michie Company, Law Publishers, Charlottesville, Virginia ; pp. 309 - 327.
- GILES, H. ; SCHERER, K. R. ; TAYLOR, D. M. ; (1979) ; « *Speech markers in social interaction* » IN: « *Social markers in speech* » (eds.: Scherer & Giles) ; Cambridge University Press, Cambridge.
- GISH, H. ; (1990) ; « *Robust discrimination in automatic speaker identification* » ; ICASSP ; pp. 289 - 292.
- GISH, H. ; KARNOFSKY, K. ; KRASHNER, M. ; ROUCOS, S. ; SCHWARTZ, R. ; WOLF, J. ; (1985) ; « *Investigation of text-independent speaker identification over telephone channels* » ; ICASSP ; pp. 379 - 382.
- GISH, H. ; KRASHNER, M. ; RUSSEL, W. ; WOLF, J. ; (1986) ; « *Methods and experiments for text-independent speaker recognition over telephone channels* » ; ICASSP ; pp. 865 - 868.
- GISH, H. ; SCHMIDT, M. ; (1994) ; « *Text-independent speaker identification* » ; IEEE Signal Processing Magazine ; no. October ; pp. 18 - 32.
- GOCKE, J. W. ; OLENIEWSKI, W. A. ; (1973) ; « *Voiceprint identification in the courtroom* » ; J. Forens. Sci. ; pp. 232 - 236.
- GOPALAN, K. ; MAHIL, S. S. ; (1991) ; « *Speaker identification and verification via singular value decomposition of speech parameters* » ; Midwest Symposium on Circuits and Systems, IEEE ; vol. 2 ; pp. 725-728.
- GORBAN, I. I. ; GORBAN, N. I. ; KLIMENKO, A. V. ; (1999) ; « *Crime-detection automatic verification and identification (CASVI) system* » ; J. Acoustic. Soc. Am. ; vol. 105 ; no. 2 ; pp. 1353.
- GROSJEAN, F. ; DESCHAMPS, A. ; (1972) ; « *Analyse des variables temporelles du français spontané* » ; *Phonetica* ; vol. 26 ; pp. 129 - 156.
- GROSJEAN, F. ; (1995) ; « *Cours de phonétique acoustique* » ; Faculté des lettres et des sciences humaines, Université de Neuchâtel.
- GRUBER, J. S. ; POZA, F. ; (1995) ; « *Voicegram identification evidence* » ; American Jurisprudence Trials, Lawyers Cooperative Publishing ; vol. 54 .
- GUBRYNOWICZ, R. ; (1973) ; « *Application of a statistical spectrum analysis to automatic voice identification* », IN: « *Speech analysis and synthesis vol. 3* » (ed.: Jassem, W.) ; Polish Academy of Sciences, Warsaw ; pp. 171 - 180.

- GUELPA, B. ; SCHAAD, B. ; (1998) ; « Georges Zecchin, mes 200 heures face à Mikhaïlov » ; L'Hebdo ; n° 51, 17 décembre.
- GUNTER, C. ; MANNING, W. ; (1982) ; « Listener estimation speaker height and weight in unfiltered and altered conditions » ; J. Phonet. ; vol. 10 ; pp. 251 - 257.
- GUYTON, A. C. ; (1984) ; « 39. La ventilation pulmonaire » IN: « Traité de physiologie médicale » ; Doin Editeurs, 8 Place de l'Odéon, 75006 Paris ; pp. 468 - 479.
- HABERSBRUNNER, H. ; SEBALD, O. ; HANTSCH, H. ; (1968) ; « Zur Personenfeststellung mittels Stimmen- und Sprachanalyse » ; Archiv für Kriminologie ; pp. 3 - 9.
- HAIR, G. D. ; REKIETA, T. W. ; (1972) ; « Automatic speaker verification using phoneme spectra » ; J. Acoust. Soc. Am. ; vol. 51 ; p. 131 (A).
- HAMMERSLEY, R. ; READ, J. D. ; (1983) ; « Testing witnesses' voice recognition : Some practical recommendations » ; J. Forensic Sci. Soc. ; vol. 23 ; pp. 203 - 208.
- HAMMERSLEY, R. ; READ, J. D. ; (1985) ; « The effect of participation in a conversation on recognition and identification of the speakers' voices » ; Law and Human Behaviour ; vol. 9 ; no. 1 ; pp. 71 - 81.
- HARTMANN, D. ; (1979) ; « The perceptual identity and characteristics of aging in normal male adult speakers » ; Journal of Communication Disorders ; vol. 12 ; pp. 53 - 61.
- HARTMANN, D. ; DANHAUER, J. ; (1976) ; « Perceptual features speech in four perceived age decades » ; J. Acoustic. Soc. Am. ; vol. 59 ; pp. 713 - 715.
- HATON, J. P. ; (1994) ; « Problems and solutions for noisy speech recognition » ; Journal de Physique IV ; vol. 4 ; no. mai ; pp. C5-439 - C5-448.
- HAYANO, T. ; (1999) ; « Les murs nippons vont avoir des oreilles » ; « Asahi Shimbun, Tokyo » IN : « Le Courier International , Paris», n° 451, 24 - 30 juin, p. 28.
- HAZEN, B. ; (1973) ; « Effects of different phonetic contexts on spectrographic speaker identification » ; J. Acoust. Soc. Am. ; vol. 54 ; pp. 650 - 660.
- HECKER, M. H. L. ; (1971) ; « Speaker recognition: an interpretive survey of the literature » ; ASHA Monographs ; vol. 16.
- HECKER, M. H. L. ; STEVENS, K. ; VON BISMARCK, G. ; WILLIAMS, C. ; (1968) ; « Manifestations of task-induced stress in the acoustic speech signal » ; J. Acoustic. Soc. Am. ; vol. 44 ; pp. 993 - 1001.
- HENNEBERT, J. ; (1998) ; « Hidden Markov models and artificial neural networks for speech and speaker recognition », thèse de doctorat n° 1860, École Polytechnique Fédérale de Lausanne.
- HENNESSY, J. J. ; ROMIG, C. H. A. ; (1971A) ; « A review of the experiments involving voiceprint identification » ; J. Forensic Sci. ; vol. 16 ; no. 2 ; pp. 183 - 198.
- HENNESSY, J. J. ; ROMIG, C. H. A. ; (1971B) ; « Sound, speech, phonetics and voiceprint identification » ; J. Forensic Sci. ; vol. 16 ; no. 4 ; pp. 438 - 454.
- HERMANSKY, H. ; (1990) ; « Perceptual linear predictive (PLP) analysis of speech » ; J. Acoust. Soc. Am. ; pp. 1738 - 1752.

- HERMAN, H. ; MORGAN, N. ; (1994) ; « RASTA Processing of speech » ; IEEE Trans. ASSP ; no. 2 ; pp. 578 - 589.
- HERZOG, H. ; (1933) ; « Stimme und Persönlichkeit » ; J. Psychol. ; vol. 130 ; pp. 300 - 379.
- HILLER, S. ; LAVER, J. ; MACKENZIE, J. ; (1984) ; « Duational aspects of long-term measurements of fundamental frequency perturbations in connected speech » ; Edinburgh University Department of Linguistics, Work in Progress ; no. 17 ; pp. 59 - 76.
- HINTZMAN, D. L. ; BLOCK, R. A. ; INSKEEP, N. R. ; (1972) ; « Memory for mode input » ; Journal of Verbal Learning and Verbal Behaviour ; vol. 11 ; pp. 741 - 749.
- HIRSON, A. ; DUCKWORTH, M. ; (1993) ; « Glottal fry and voice disguise: A case study in forensic phonetics » ; Journal of Biomedical Engineering ; vol. 15 ; no. 3 ; pp. 193 - 200.
- HOLLIEN, H. ; (1977) ; « Status report on voiceprint identification in the United States » ; Proceedings of the International Conference on Crime Countermeasures, Science and Engineering, Oxford ; July 25th - 29th.
- HOLLIEN, H. ; (1990) ; « The acoustics of crime » ; Plenum Press, New York.
- HOLLIEN, H. ; (1995) ; « The Future in Speaker Identification: A Model » ; Proceedings of the XIIth International Congress of Phonetic Sciences, Stockholm ; vol. 3 ; pp. 138 - 145.
- HOLLIEN, H. ; JIANG, M. ; (1998) ; « The challenge of effective speaker identification » ; Proceedings of RLA2C Workshop: « Speaker Recognition and its Commercial and Forensic Applications » ; pp. 2 - 10.
- HOLLIEN, H. ; MAJEWSKI, W. ; (1977) ; « Speaker identification by long-term spectra under normal and distorted speech conditions » ; J. Acoust. Soc. Am. ; vol. 62 ; no. 4 ; pp. 975 - 980.
- HOLLIEN, H. ; MAJEWSKI, W., DOHERTY, E. T. ; (1982) ; « Perceptual identification of voices under normal, stress and disguised speaking conditions » ; J. Phonetics ; no. 10 ; pp. 139 - 148.
- HOLLIEN, H. ; MARTIN, C. A. ; (1996) ; « Conducting research on the effects of intoxication on speech » ; Forensic Linguistics ; vol. 3 ; no. 1 ; pp. 107 - 129.
- HOLLIEN, H. ; MCGLONE, R. E. ; (1976) ; « The effects of disguise on voiceprint identification » ; Journal of Criminal Defense ; no. 2 ; pp. 117 - 130.
- HOLLIEN, H. ; SHIPP, F. T. ; (1972) ; « Speaking fundamental frequency and chronological age in males » ; J. Speech Hearing Res. ; vol. 15 ; pp. 155 - 159.
- HOMAYOUNPOUR, M. M. ; CHOLLET, G. ; (1995) ; « A study of intra- and inter-speaker variability in voices of twins for speaker verification » ; Proceedings of the XIIth International Congress of Phonetic Sciences, Stockholm ; vol. 3 ; pp. 298 - 301.
- HOMAYOUNPOUR, M. M. ; GOLDMAN, J. P. ; CHOLLET, G. ; (1993) ; « Machine vs. human speaker verification » ; IAFP Conference, Trier.
- HORII, Y. ; (1975) ; « Some statistical characteristics of voice fundamental frequency » ; J. Speech Hearing Res. ; vol. 18 ; pp. 192 - 201.
- HORII, Y. ; RYAN, W. ; (1981) ; « Fundamental frequency characteristics and perceived age of adult male speakers » ; Folia Phoniatica ; vol. 33 ; pp. 227 - 233.

- HUNT, A. K. ; (1991) ; « *New commercial applications of telephone-network-based speech recognition and speaker verification* » ; Eurospeech 91 ; pp. 431 - 433.
- HUNT, M. ; (1983) ; « *Further experiments in text-independent speaker recognition over communications channels* » ; ICASSP ; pp. 563 - 566.
- HUNTLEY, R. ; HOLLIEN, H. ; SHIPP, T. ; (1987) ; « *Influences of listener characteristics on perceived age estimations* » ; J. Voice ; vol. 1 ; pp. 49 - 52.
- INGEMAN, F. ; (1968) ; « *Identification of the speaker's sex from voiceless fricatives* » ; J. Acoustic. Soc. Am. ; vol. 44 ; pp. 1142 - 1144.
- INGRAM, J. C. L. ; (1995) ; « *Formant trajectories for speaker identification: where they work, and where they don't* » ; Paper to the International Association of Forensic Linguistics Conference, Armidale.
- INGRAM, J. C. L. ; PRANDOLINI, R. ; ONG, S. ; (1996) ; « *Formant trajectories as indices of phonetic variation for speaker identification* » ; Forensic Linguistics ; vol. 3 ; no. 1 ; pp. 129 - 145.
- JANKOWSKI, C. R. ; QUATIERI, T. F. ; REYNOLDS, D. A. ; (1994) ; « *Formant AM-FM for speaker identification* » ; Proc. IEEE SP Int. Symp. Time Freq. Time Scale Anal. ; pp. 608 - 611.
- JASSEM, W. ; STEFFEN-BATOG, M. ; CZAJKA, S. ; (1973) ; « *Statistical characteristics of short-term average F_0 distributions as personal voice features* » IN: « *Speech analysis and synthesis* » (ed.: Jassem, W.) ; Polish Academy of Sciences: Warsaw ; vol. 3 ; pp. 209 - 225.
- JAYANT, N. ; (1992) ; « *High-quality coding of telephone speech* » in « *Advances in speech signal processing* » (eds. Furui S., Sondhi M.) ; Dekker, New York, USA ; pp. 85 - 108.
- JONES, W. R. ; (1973A) ; « *Danger - Voiceprint ahead* » ; Amer. crim. Law Rev. ; vol. 11 ; no. 3 ; pp. 549 - 573.
- JONES, W. R. ; (1973B) ; « *Evidence vel non. The nonsense of voiceprint identification* » ; Kentucky Law J. ; vol. 62 ; no. 2 ; pp. 301 - 326.
- KAISER, L. ; (1939 - 1944) ; « *Biological and statistical research concerning the speech of 216 Dutch students I - V* » ; Archives néerlandaises de phonétique expérimentale ; no. 15, 16, 17, 18.
- KACZMAREK, Z. ; KRZYSZKO, M. ; (1973) ; « *An attempt to use Anderson and Bahadur's separating hyperplane to identify a population among many normal populations* », IN: « *Speech analysis and synthesis* » (ed.: Jassem, W.) ; Polish Academy of sciences: Warsaw ; vol. 3 ; pp. 159 - 169.
- KAO, Y. H. ; BARRAS, J. S. ; RAJASEKARAN, P. K. ; (1993) « *Robustness study of free-text speaker identification and verification* » ; ICASSP ; pp. II-379-II-382.
- KAYE, D. H. ; (1979) ; « *The Laws of Probability and the Law of the Land* » ; The University of Chicago Law Review ; no. 47 ; pp. 34 - 56.
- KELLER, E. ; (1994) ; « *Signalize™, analyse du signal pour la parole et le son, manuel d'utilisation* » ; Network Technology Corporation, Charlestown.
- KERSTA, L. G. ; (1962A) ; « *Voiceprint identification* » ; Nature ; no. 4861 ; pp. 1253-1257.
- KERSTA, L. G. ; (1962B) ; « *Voiceprint identification* » ; J. Acoust. Soc. Am. ; vol. 34 ; p. 725 (A).
- KERSTA, L. G. ; (1973) ; « *L'identification des voix* » ; RIPC ; pp. 9 - 15.

- KERSTA, L. G. ; NASH, E. W. ; (1973) ; « *Voiceprint identification. Observations I and II* » ; Int. Crim. Police Rev. ; vol. 28 ; no. 264 ; pp. 9 - 15.
- KLEVANS, L. ; RODMAN, R. D. ; (1997) ; « *Voice recognition* » ; Artech House, Boston, MA.
- KOENIG, B. ; (1980) ; « *Speaker identification: Three methods - listening, machine and aural-visual* » ; FBI Law Enforcement Bulletin ; no. Jan ; pp. 1 - 4.
- KOENIG, B. E. ; (1986A) ; « *Spectrographic voice identification: a forensic survey* » ; J. Acoust. Soc. Am. ; vol. 79 ; no. 6 ; pp. 2088 - 2090.
- KOENIG, B. E. ; (1986B) ; « *Spectrographic voice identification* » ; Crime Laboratory Digest ; vol. 13 ; no. 4 ; pp. 105 - 118.
- KOENIG, B. E. ; RITENOUR, D. V. ; KOHUS, B. A. ; SAVOY KELLY, A. ; (1987) ; « *Reply to: 'Some fundamental considerations regarding voice identification'* » ; J. Acoustic. Soc. Am. ; vol. 82 ; no. 2 ; pp. 688 - 689.
- KOIKE, Y. ; (1973) ; « *Application of some acoustic measures for the evaluation of laryngeal dysfunction* » ; Studia Phonologica (Kyoto University) ; vol. 7 ; pp. 17 - 23.
- KONDOZ, A. M. ; (1994) ; « *Coding for low bit rate communication systems* » ; John Wiley & Sons, New York, USA.
- KOPP, G. A. ; GREEN, H. C. ; (1946) ; « *Basic phonetic principles of visible speech* » ; J. Acoust. Soc. Am. ; no. 18 ; pp. 74 - 90.
- KÖSTER, J. P. ; (1987) ; « *Auditive Sprechererkennung bei Experten und Naiven*, IN: « *Festschrift für H. Wrängler* » (ed.: Weiss, R.) ; Buske, Hamburg ; pp. 171 - 180.
- KOVAL, S. ; ILYINA, O. ; KHITINA, M. ; (1998A) ; « *Practice of Usage of Auditive and Linguistic Features for Forensic Speaker Identification* » ; Proceedings of the 8th COST 250 workshop, Ankara: « *Speaker identification by man and by machine: Directions for forensic applications* » ; pp. 23 - 29.
- KOVAL, S. ; KAGANOV, A. ; KITHROY, M. ; (1998B) ; « *The Chart of the Standard Expert Actions and Decision Making Principles of Forensic Speaker Identification* » ; Proceedings of the 8th COST 250 workshop, Ankara: « *Speaker identification by man and by machine: Directions for forensic applications* » ; pp. 62 - 66.
- KRASHNER, M. ; WOLF, J. ; KARNOFSKY, K. ; SCHWARTZ, R. ; ROUCOS, S. ; GISH, H. ; (1984) ; « *Investigation of text-independent speaker identification techniques under conditions of variable data* » ; ICASSP 84 ; pp. 18b.5. 1-4.
- KRAUSE, H. J. ; (1976) ; « *Possibilities of identification by voice and limits* » ; Arch. Krim. ; vol. 157 ; no. 5 & 6 ; pp. 154 - 164.
- KREBSER, U. G. ; (1993) ; « *Frequenz-Handbuch: der mobilen und festen Funkdienste der Schweiz* » ; Poly-Verlag, Bassersdorf.
- KRETSCHMER, E. ; (1922) ; « *Körperbau und Charakter : Untersuchungen zum Konstitutionsproblem und zur Lehre von den Temperamenten* » ; J. Springer, Berlin.
- KRZYSZKO, M. ; JASSEM, W. ; FRACKOWIAK-RICHTER, L. ; (1973) ; « *Statistical discrimination functions and their applications to the problem of voice identification* », IN: « *Speech analysis and synthesis* » (ed.: Jassem, W.) ; Polish Academy of sciences, Warsaw ; vol. 3 ; pp. 144 - 157.

- KÜNZEL, H. J. ; (1987) ; « *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung* » ; Kriminalistik Verlag, Heidelberg.
- KÜNZEL, H. J. ; (1989) ; « *How well does average fundamental frequency correlate with speaker height and weight* » ; *Phonetica* ; no. 46 ; p. 117.
- KÜNZEL, H. J. ; (1994A) ; « *Current approaches to forensic speaker recognition* » ; Proceedings of ESCA Workshop on automatic speaker recognition, identification and verification ; pp. 135 - 141.
- KÜNZEL, H. J. ; (1994B) ; « *On the problem of speaker identification by victims and witnesses* » ; *Forensic Linguistics* ; vol. 1 ; no. 1 ; pp. 45 - 58.
- KÜNZEL, H. J. ; MASTHOFF, H. R. ; KÖSTER, J. P. ; (1995) ; « *The relation between speech tempo, loudness, and fundamental frequency: an important issue in forensic speaker recognition* » ; *Science & Justice* ; vol. 35 ; no. 4 ; pp. 291 - 295.
- KWAN, Q. Y. ; (1977) ; « *Inference of Identity of Source* » ; Ph. D. Thesis, University of California, Berkeley, CA, USA.
- LABOV, W. ; (1972) ; « *Sociolinguistic patterns* » ; University of Pennsylvania Press, Philadelphia.
- LADD, D. R. ; SILVERMAN, K. E. A. ; TOLKMITT, F. ; BERGMANN, G. ; SCHERER, K. R. ; (1985) ; « *Evidence for the independent function of intonation contour type, voice quality and F₀ range in signaling speaker affect* » ; *J. Acoustic. Soc. Am.* ; vol. 78 ; pp. 435 - 444.
- LADEFOGED, P. N. ; (1962) ; « *Elements of acoustic phonetics* » ; The University of Chicago Press, Chicago, USA.
- LADEFOGED, P. ; V ANDERSLICE, R. ; (1967) ; « *The voiceprint mystique* » ; Working papers in phonetics, University of California, Los Angeles ; no. 7, November.
- LAMEL, L. ; GAUVAIN, J. L. ; (1998) ; « *Speaker verification over the telephone* » ; Proceedings of RLA2C Workshop : « *Speaker recognition and its commercial and forensic applications* » ; pp. 76 - 79.
- LANGANEY, A. ; (1992) ; « *Les races existent-elles ?* » ; *Sciences et avenir* ; no. 540 ; pp. 45 - 50.
- LARIVIERE, C. L. ; (1971) ; « *Contributions of fundamental frequency and formant frequencies to speaker identification?* » ; *Phonetica* ; no. 31 ; pp. 185 - 197.
- LASHBROOK, W. B. ; (1972) ; « *An examination of conditional variations for voice identification trials* », IN: « *Voice Identification Research* » ; U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice ; pp. 119 - 136.
- LASS, N. J. ; BEVERLY, A. S. ; NICOSIA, D. K. ; SIMPSON, L. A. ; (1978) ; « *An investigation of speaker height and weight identification by means of direct estimations* » ; *J. Phonet.* ; vol. 6 ; pp. 69 - 76.
- LASS, N. J. ; HUGHES, K. R. ; BOWYER, M. D. ; (1976) ; « *Speaker sex identification from voiced, whispered and filtered isolated vowels* » ; *J. Acoustic. Soc. Am.* ; vol. 59 ; pp. 675 - 678.
- LASS, N. J. ; KELLY, D. T. ; CUNNINGHAM, C. M. ; (1980A) ; « *A comparative study of speaker height and weight identification from voiced and whispered speech* » ; *W. Va. Univ. J. Phonet.* ; vol. 12.
- LASS, N. ; PHILIPS, J. K. ; BRUCHEY, C. A. ; (1980B) ; « *The effect of filtered speech on speaker height and weight identification* » ; *J. Phonet.* ; vol. 8 ; pp. 90 - 100.

- LASS, N. J. ; ALMERINO, C. A. ; JORDAN, L. F. ; WALSH, J. N. ; (1980C) ; « *The effect of filtered speech on speaker race and sex identification* » ; J. Phonet. ; vol. 8 ; pp. 101 - 112.
- LAVER, J. ; (1980) ; « *The phonetic description of voice quality* » ; Cambridge University Press.
- LEMPERT, R. O. ; (1995) ; « *The honest scientist's guide to DNA evidence* » ; IN: « *Human identification: the use of DNA markers* » (eds Weir, B. S.) ; Kluwer Academic Publishers, Dordrecht ; Vol. 4. ; pp. 119-124.
- LEVIN, H. ; LORD, W. ; (1975) ; « *Speech pitch frequency as emotional state indicator* » ; IEEE Trans. SMC ; vol. 5 ; no. 2 ; p. 259.
- LEWIS, S. R. ; (1984) ; « *Philosophy of Speaker Identification* » ; Police Applications of Speech and Tape Recording Analysis: Proceedings of Acoustics ; vol. 6 ; no. 1 ; pp. 69 - 77.
- LI, K. ; HUGHES, G. ; (1974) ; « *Talker differences as they appear in correlation matrices of continuous speech spectra* » ; J. Acoust. Soc. Am. ; vol. 55 ; pp. 833 - 837.
- LI, K. ; DAMMANN, J. E. ; CHAPMAN, W. D. ; (1966) ; « *Experimental studies in speaker verification using an adaptative system* » ; J. Acoust. Soc. Am. ; vol. 40 ; pp. 966 -978.
- LIGHT, L. L. ; STANBURY, C. ; RUBINS, C. ; LINDE, S. ; (1973) ; « *Memory for modality of presentation: Within-modality discrimination* » ; Journal of Applied Psychology ; no. 1 ; pp. 395 - 400.
- LIN, W. C. ; PILLAY, S. K. ; (1976) ; « *Feature evaluation and selection for an on-line adaptative speaker verification system* » ; IEEE Conf. ASSP ; pp. 734 - 737.
- LINDLEY, D. V. ; (1977) ; « *Probability and the Law* » ; The Statistician ; vol. 26 ; no. 3 ; pp. 203 - 220.
- LIPEIKA, A. ; LIPEIKIENE, J. ; (1997) ; « *Speaker identification methods based on pseudostationary segments of voiced sounds* » ; Informatica ; vol. 7 ; no. 4 ; pp. 469 - 484.
- LOCARD, E. ; (1909) ; « *L'identification des récidivistes* » ; A. Maloine, 25 - 27 Rue de l'école de médecine, Paris.
- LOCARD, E. ; (1932) ; « *Le signalement* » IN: « *Les preuves de l'identité* » ; Joannès Desvignes et ses fils, libraires - éditeurs, 36 à 42, Passage de l'Hôtel-Dieu, Lyon ; pp. 201 - 203.
- LOCARD, E. ; (1959) ; « *Les faux en écriture et leur expertise* » ; Payot, Paris, 106, Boulevard Saint-Germain, pp. 282 - 293.
- LOEVINGER, L. ; (1995) ; « *Science as evidence* » ; Jurimetrics, Journal of law, science and technology ; vol. 35 ; no. 2 ; pp. 153 - 190.
- LUCK, J. E. ; (1969) ; « *Automatic speaker verification using cepstral measurements* » ; J. Acoust. Soc. Am. ; vol. 46 ; pp. 1026 - 1032.
- LUMMIS, R. ; (1973) ; « *Speaker verification by computer using speech intensity for temporal registration* » ; IEEE Trans. Audio & Electroac. ; vol. AU-21 ; pp. 80 - 89.
- MAJEWSKI, W. ; ZALEWSKI, J. ; HOLLIEN, H. ; (1979) ; « *Some remarks on different speaker identification techniques* » ; IN : « *Current Issues in linguistic Theory* » (eds. Hollien H. & P.) ; John Benjamins B. V., Amsterdam ; vol. 9, pp. 829 - 835.
- MAKHOUL, J. ; ROUCOS, S. ; GISH, H. ; (1985) ; « *Vector quantization in speech coding* » ; Proc. IEEE ; vol. 73 ; no. 11 ; pp. 1551 - 1588.

- MAMMONE, R. J. ; ZHANG, X. ; RAMACHANDRAN R. P. ; (1996) ; « *Robust speaker recognition: A feature-based approach* » ; IEEE Signal Processing Magazine ; pp. 58 - 71.
- MAMOUX J. P. ; (1971) ; « *Identification de la voix humaine* » ; Médecine légale et dommage corporel ; no. 4 ; pp. 35 - 38.
- MANDELBROT, B. ; (1983) ; « *The fractal geometry of nature* » ; W.H. Freeman, New York , USA.
- MARESCAL, F. ; (1999) ; « *The forensic speaker recognition method used in the French Gendarmerie* » ; European Union Symposium of Forensic Science, Wiesbaden.
- MARTIN, E. P. ; (1967) ; « *Zur Frage des Beweiswertes von Tonbandaufnahmen im Strafprozess* » ; Kriminalistik ; pp. 511 - 518.
- MASTHOFF, H. ; (1996) ; « *A report on a voice disguise experiment* » ; Forensic Linguistics ; vol. 3 ; no. 1 ; pp. 160 - 168.
- MATALON, B. ; (1967) ; « *Epistémologie des Probabilités* » IN: « *Encyclopédie de la Pléiade: Logique et connaissance scientifique* » (éd. Piaget, J.) ; Gallimard, Dijon, France ; vol. XXII ; pp. 526 - 553.
- MATHYER, J. ; (1990) ; « *Lettre à Mesdames et Messieurs les magistrats de l'ordre judiciaire* » ; Revue Int. Crim. Pol. Tech. ; vol. XLIII ; no. 1/90 ; pp. 98 - 100.
- MATSUI, T. ; FURUI, S. ; (1991) ; « *A text-independent speaker recognition method robust against utterance variations* » ; ICASSP ; vol. 1 ; pp. 377 - 380.
- MATSUI, T. ; FURUI, S. ; (1992) ; « *Comparison of text-independent speaker recognition methods using vector quantization distortion and discrete and continuous HMMs* » ; ICASSP ; vol. A ; pp. II-157 - II-160.
- MCDADE, T. ; (1968) ; « *The voiceprint* » ; The Criminologist ; vol. 3, no. 7 ; pp. 52 -70.
- MCGEHEE, F. ; (1937) ; « *The reliability of the identification of human voice* » ; J. Gen. Psychol. ; vol. 17 ; pp. 249 - 271.
- MCGEHEE, F. ; (1944) ; « *An experimental study of voice recognition* » ; J. Gen. Psychol. ; vol. 31 ; pp. 53 - 65.
- MCGONEGAL, C. ; ROSENBERG, A. ; RABINER, L. ; (1978) ; « *Speaker verification by human listeners over several speech transmission systems* » ; Bell Sys. Tech. Journal ; vol. 57 ; no. 8 ; pp. 2887 - 2900.
- MELLA, O. ; (1992) ; « *Pertinence des trois premiers formants des voyelles dans l'identification* » ; 19èmes JEP ; pp. 549 - 555.
- MELLA, O. ; (1994) ; « *Extraction of formants of oral vowels and critical analysis for speaker characterization* » ; Proceedings of ESCA Workshop on automatic speaker recognition, identification and verification ; pp. 193 - 196.
- MELVIN, C. ; NAKASONE, H. ; TOSI, O. ; (1988) ; « *More fundamental considerations regarding voice identification* » ; J. Acoust. Soc. Am. ; vol. 84 ; no. 5 ; pp. 1943 - 1944.
- MERKEL, F. ; (1902) ; « *Atmungsorgane* » IN: « *Darmsystem* » (hrsg. von Bardeleben, K.) ; Jena: Verlag von Gustav Fischer ; pp. 39 - 40.
- MERMINOD, Y. ; (1992) ; « *Expressions et proverbes latins, adages juridiques* » ; Ides & Calendes ; Neuchâtel.

- MERTZ, N. J. ; KIMMEL, K. L. ; (1978) ; « *The effect of temporal speech alterations on speaker race and sex identifications* » ; Language and Speech ; vol. 21 ; pp. 279 - 290.
- MEUWLY, D. ; (1999) ; « *L'ordonnance sur le service de surveillance de la correspondance postale et des télécommunications du 1.12.1997 : Une loi en retard d'une guerre technologique ?* » IN: « *Le statut des télécommunications en mutation* », Editions Universitaires, Fribourg.
- MEUWLY, D. ; DRYGAJLO, A. ; (1997) ; « *Likelihood ratios for automatic speaker recognition in forensic applications* », poster présenté lors de la conférence annuelle de l'IAFP à Edinbourg.
- MEUWLY, D. ; EL MALIKI, M. ; DRYGAJLO, A. ; (1998) ; « *Forensic Speaker Recognition Using Gaussian Mixture Models and a Bayesian Framework* » ; Proceedings of the 8th COST 250 workshop, Ankara: « *Speaker identification by man and by machine: Directions for forensic applications* » ; pp. 52 - 55.
- MEUWLY, D. ; (2000) « *Voice Analysis* » IN : « *Encyclopedia of forensic science* » (eds. Siegel, J. ; Saukko, P. and Knupfer, G.) ; Academic Press Ltd, London, UK ; pp. 1413 - 1423.
- MOENSSENS, A. A. ; INBAU, F. E. ; STARRS, J. E. ; (1986) ; « *Spectrographic voice recognition* » IN: « *Scientific evidence in criminal cases* » ; 3rd ed., The foundation Press, Inc., Mineola, New York, USA ; pp. 653 - 677.
- MOON, T. K. ; (1996) ; « *The Expectation-maximisation algorithm* » ; Proc. IEEE ; pp. 47 - 60.
- MOSES, J. P. ; (1941) ; « *Theories regarding the relation of constitution and character through the voice* » ; Psychol. Bull. ; vol. 38 ; pp. 746.
- NAIK, J. M. ; NETSCH, L. P. ; DODDINGTON, G. R. ; (1989) ; « *Speaker verification over long-distance telephone lines* » ; ICASSP ; pp. 524 - 527.
- NAKASONE, H. ; (1999) ; « *Communication personnelle* », 30 avril.
- NAKASONE, H. ; MELVIN, C. ; (1989) ; « *C.A.V.I.S.: (Computer Assisted Voice Identification System)* » ; final report ; National Institute of Justice, Grant no. 85-IJ-CX-0024.
- NASH, E. W. ; (1973) ; « *Voice identification with the aid of spectrographic analysis* » ; J. Assoc. Offic. Analyt. Chem. ; vol. 56 ; no. 4 ; pp. 944 - 946.
- NATARAJAN, M. ; CLARKE, R. V. ; JOHNSON, B. D. ; (1995) ; « *Telephones as facilitators of drug dealing in crime environments and situational prevention* » ; European Journal of Criminal Policy and Research ; vol. 03.3 ; pp. 137 - 153.
- NEIMANN, G. S. ; APPLGATE, J. A. ; (1990) ; « *Accuracy of listeners judgements of perceived age relative to chronological age in adults* » ; Folia Phoniatica ; vol. 42 ; pp. 327 - 330 .
- NOLAN, F. ; (1983) ; « *The phonetic bases of speaker recognition* » ; Cambr. Univ. Press, Cambridge, UK.
- NOLAN, F. ; (1990) ; « *The limitations of auditory-phonetic speaker identification*, IN: 'Texte zu Theorie und Praxis forensischer Linguistik' » (ed: Kniffka, H.) » ; M. Niemeyer, Tübingen ; pp. 457 - 479.
- NOLAN, F. ; (1991) ; « *Forensic phonetics* » ; Journal of Linguistics ; vol. 27 ; pp. 483 - 493.
- NOLAN, F. ; (1992) ; « *Code of practice* » ; Journal of the International Phonetic Association ; vol. 1 & 2 ; pp. 80 - 81.

- NOLAN, F. ; (1995) ; « *Can the definition of each speaker be expected to come from the laboratory in the next decade* » ; Proceedings of the XIIth International Congress of Phonetic Sciences, Stockholm ; vol. 3 ; pp. 130 - 137.
- NOLL, P. ; (1975) ; « *Technische Methoden zur Überwachung verdächtiger Personen im Strafverfahren* » ; Revue Pénale Suisse ; vol. 91 ; pp. 45 - 73.
- O'SHAUGNESSY, D. ; (1986) ; « *Speaker recognition* » ; IEEE ASSP Magazine ; vol. 3 ; no. 4 ; pp. 4 -17.
- O'SHAUGNESSY, D. ; (1987) ; « *Speaker recognition* » IN: « *Digital speech processing, synthesis, and recognition* » ; Addison-Wesley Publishing Company, New York.
- OPENSHAW, J. P. ; SUN, Z. P. ; MASON, J. S. ; (1993) ; « *Comparison of composite features under degraded speech in speaker recognition* » ; ICASSP ; vol. 2 ; pp. II-371 - II-374.
- ORMEZZANO, Y. ; ROCH, J.-B. ; (1991) ; « *Analyse vocale immédiate normalisée* » ; Bulletin d'audiophonologie ; vol. 21 ; no. 4 ; pp. 399 - 452.
- ORTEGA-GARCIA, J. ; GONZALEZ-RODRIGUEZ, J. ; MARRERO-AGUIAR, V. ; DIAZ-GOMEZ, J. J. ; GARCIA-JIMENEZ, R. ; LUCENA-MOLINA, J. ; SANCHEZ-MOLERO, J. A. G. ; (1998) ; « *Speaker verification in forensic tasks using 'AHUMADA' speech corpus* » ; Proceedings of RLA2C Workshop: « *Speaker Recognition and its Commercial and Forensic Applications* » ; pp. 141 -144.
- OTTOLENGHI, S. ; (1910) ; « *Trattato di polizia scientifica* » ; Societa Editrice Libreria, Via Kramer 4A, Milano ; pp. 272 - 276.
- PAOLONI, A. ; (1999) ; « *Communication personnelle* » ; 6 mai.
- PAOLONI, A. ; FALCONE, M. ; BIMBOT, F. ; CHOLLET, G. ; (1994) ; « *Outline a comprehensive assessment methodology for speaker recognition task in forensic application* » ; Annual Meeting of The International Association for Forensic Linguistics, Cardiff.
- PAPAMICHALIS, P. E. ; DODDINGTON, G. R. ; (1984) ; « *A speaker recognizability test* » ; ICASSP 84 ; no. 18B.6. ; pp. 1 - 4.
- PAUL, J. E. ; RABINOWITZ, A. S. ; RIGANATI, J. P. ; RICHARDSON, J. M. ; (1975) ; « *Semi-automatic speaker identification system (SASIS) - Analytical studies* » ; Final Report, Rockwell International Report N° C74-11841501.
- PAWLEWSKI, M. ; DOWNEY, S. N. ; (1996) ; « *Channel effects in speaker recognition* » ; BT Technology Journal ; no. January.
- PERKELL, J. S. ; KLATT, D. S. ; STEVENS, K. N. ; KEYERS, S. J. ; (1986) ; « *Toward a phonetic and phonological theory of redundant features* » IN: « *Invariance and variability in speech processes* » (eds: Perkell, J. S. & Klatt, D. H.) ; L. Erlbaum , London ; pp. 426 - 449.
- PIQUEREZ, G. ; (1994) ; « *Précis de procédure pénale suisse* » ; Payot, Lausanne.
- PISONI, D. B. ; LUCE, P. A. ; (1987) ; « *Acoustic-phonetic representations in word recognition* » ; Cognition ; pp. 21 - 52.
- POLLACK, I. ; PICKETT, J. M. ; SUMBY, W. H. ; (1954) ; « *On the identification of speakers by voice* » ; J. Acoustic. Soc. Am. ; vol. 26 ; pp. 403 - 406.

- POPPER, K. R. ; (1973 (édition originale 1935)) ; « *La falsifiabilité* », IN: « *La logique de la découverte scientifique* » ; Payot, Lausanne, Suisse ; pp. 77 - 91.
- POPPER, K. R. ; (1988 (premières publications 1944/45)) ; « *Misère de l'historicisme* » ; Pocket, Paris.
- POTTER, R. K. ; (1946) ; « *Introduction to technical discussions of sound portrayal* » ; J. Acoust. Soc. Am. ; no. 18 ; pp. 1 - 3.
- POTTER, R. K. ; KOPP, K. G. ; GREEN, H. C. ; (1947) ; « *Visible speech* » ; D. van Nostrand Co, NY.
- POZA, F. T. ; (1999) ; « *Communication personnelle* », 14 avril.
- PRESTI, A. ; (1966) ; « *High Speed Sound Spectrograph* » ; J. Acoust. Soc. Am. ; vol. 40 ; pp. 628 - 634.
- PRUZANSKY, S. ; (1963) ; « *Pattern-matching procedure for automatic talker recognition* » ; J. Acoust. Soc. Am. ; vol. 35 ; pp. 354 - 358.
- PRUZANSKY, S. ; MATHEWS, V. ; (1964) ; « *Talker-recognition procedure based on analysis of variance* » ; J. Acoust. Soc. Am. ; vol. 36 ; pp. 2041 - 2047.
- PRZYBOCKI, M. ; MARTIN, A. F. ; (1998) ; « *NIST speaker recognition evaluation - 1997* » ; Proceedings of RLA2C Workshop: « *Speaker Recognition and its Commercial and Forensic Applications* » ; pp. 120 - 124.
- PTACEK, P. H. ; SANDER, E. K. ; (1966) ; « *Age recognition from voice* » ; J. Speech Hearing Res. ; vol. 9 ; pp. 273 - 277.
- PTACEK, P. H. ; SANDER, E. K. ; MALONEY, W. H. ; (1966) ; « *Phonatory and related changes with advanced age* » ; J. Speech Hearing Res. ; vol. 9 ; pp. 353 - 360.
- RABINER, L. R. ; JUANG, B. H. ; (1993) ; « *Fundamental of speech recognition* » ; PTR Prentice-Hall.
- RABINER, L. ; SCHAFER, R. ; (1978) ; « *Digital processing of speech signals* » ; Englewood cliffs, NJ, Prentice Hall.
- RAMACHANDRAN, R. P. ; ZILOVIC, M. S. ; MAMMONE, R. J. ; (1995) ; « *Comparative study of robust linear predictive analysis methods with applications to speaker identification* » ; IEEE Trans. ASSP ; vol. 3 ; no. 2 ; pp. 117 - 125.
- RAMIG, L. A. ; RINGEL, R. L. ; (1983) ; « *Effects of physiological aging on selected acoustic characteristics of voice* » ; J. Speech Hearing Res. ; vol. 26 ; pp. 22 - 30.
- RAMISHVILI, G. S. ; (1966) ; « *Automatic voice recognition* » ; Eng. Cyber. ; vol. 5 ; pp. 84 - 90.
- REDNER, R. A. ; WALKER, H. F. ; (1984) ; « *Mixture densities, maximum likelihood and the EM algorithm* » ; SIAM Review ; vol. 26 ; no. 2 ; pp. 195 - 239.
- REICH, A. ; DUKE, J. ; (1979) ; « *Effects of selected vocal disguises upon speaker identification by listening* » ; J. Acoustic. Soc. Am. ; vol. 66 ; pp. 1023 - 1028.
- REICH, A. ; MOLL, K. ; CURTIS, J. ; (1976) ; « *Effects of selected vocal disguises upon spectrographic speaker identification* » ; J. Acoust. Soc. Am. ; vol. 60 ; pp. 919 - 925.
- REISS, A. R. ; (1907) ; « *Un code télégraphique du portrait parlé* » ; A. Maloine, 25 - 27 Rue de l'école de médecine, Paris ; pp. 17 - 18.
- REYNOLDS, D. A. ; (1992) ; « *A gaussian mixture modeling approach to text-independent speaker identification* » ; Ph. D. thesis, Georgia Institute of Technology, Atlanta, USA.

- REYNOLDS, D. A. ; (1994) ; « *Speaker identification and verification using gaussian mixture speaker models* » ; Proceedings of ESCA Workshop on automatic speaker recognition, identification and verification ; pp. 27 - 30.
- REYNOLDS, D. A. ; (1995A) ; « *Automatic speaker recognition using gaussian mixture speaker models* » ; The Lincoln Laboratory Journal ; vol. 8 ; no. 2 ; pp. 173 - 191.
- REYNOLDS, D. A. ; (1995B) ; « *Speaker identification and verification using gaussian mixture speaker models* » ; Speech Communication ; vol. 17 ; no. 1 - 2 ; pp. 91 -108.
- REYNOLDS, D. A. ; (1996) ; « *The effects of handset variability on speaker recognition performance: experiments on the switchboard corpus* » ; IEEE Trans. ASSP ; pp. 113 - 116.
- REYNOLDS, D. A. ; ROSE, R. C. ; (1995) ; « *Robust text-independent speaker identification using Gaussian mixture speaker models* » ; IEEE Trans. ASSP ; vol. 3 ; no. 1 ; pp. 72 - 83.
- REYNOLDS, J. C. ; WEBER, J. W. ; (1979) ; « *The admissibility of spectrographic voice identification in the state courts* » ; J. Crim. Law & Criminology ; vol. 70 ; no. 3 ; pp. 349 - 354.
- RIBAU, O. ; (1997) ; « *La recherche et gestion des liens dans l'investigation criminelle: le cas particulier du cambriolage* » ; thèse de doctorat, Institut de police scientifique et de criminologie, Université de Lausanne.
- RINGEL, R. L. ; CHODZKO-ZAJKO, W. J. ; (1987) ; « *Vocal indices of biological age* » ; Journal of Voice ; vol. 1 ; pp. 31 - 37.
- ROBERTSON, B. ; VIGNAUX, G. A. ; (1995) ; « *Interpreting Evidence: Evaluating Forensic Science in the Courtroom* » ; John Wiley & Sons ; Chichester, UK.
- RODMAN, R. D. ; (1998) ; « *Speaker recognition of disguised voices: A program for research* » ; Proceedings of the 8th COST 250 workshop, Ankara: « *Speaker identification by man and by machine: Directions for forensic applications* » ; pp. 9 - 22.
- ROSE, P. ; DUNCAN, S. ; (1995) ; « *Naïve auditory identification and discrimination of similar voices by familiar listeners* » ; Forensic Linguistics ; vol. 2 ; no. 1 ; pp. 1 - 17.
- ROSE, R. C. ; FITZMAURICE, J. ; HOFSTETTER, E. M. ; REYNOLDS, D. A. ; (1991) ; « *Robust speaker identification in noisy environments using noise adaptive speaker models* » ; ICASSP ; vol. 1 ; pp. 401 - 404.
- ROSE, R. C. ; REYNOLDS, D. A. ; (1990) ; « *Text-independent speaker identification using automatic acoustic segmentation* » ; ICASSP 90 ; pp. 293 - 296.
- ROSENBERG, A. E. ; (1973) ; « *Listener performance in speaker verification tasks* » ; IEEE Trans. Audio Electroacoust. ; no. 3 ; pp. 221 - 225.
- ROSENBERG, A. E. ; (1976A) ; « *Automatic speaker verification: a review* » ; Proc. IEEE ; vol. 64 ; no. 4 ; pp. 475 - 487.
- ROSENBERG, A. E. ; (1976B) ; « *Evaluation of an automatic speaker verification system over telephone lines* » ; B. S. T. J. ; vol. 55 ; no. 6 ; pp. 723 - 743.
- ROSENBERG, A. E. ; SOONG, F. K. ; (1986) ; « *Evaluation of a vector quantization talker recognition system in text-independent and text-dependent modes* » ; ICASSP ; pp. 873 - 876.

- ROSENBERG, A. E. ; SOONG, F. K. ; (1991) ; « *Recent research in automatic speaker recognition* », IN: « *Advances in speech signal processing* » (eds: Furui S. & Sondhi M. M.) ; Marcel Decker, New York, USA ; pp. 701 - 737.
- ROTHER, H. ; (1967) ; « *Stimm-Spektrographie: Neuartiges Hilfsmittel der Kriminalistik* » ; Kriminalistik ; pp. 233 - 235.
- ROTHMAN, H. B. ; (1979) ; « *Further analysis of talkers with similar sounding voices* » IN: « *Current issues in linguistic theory* » (eds. Hollien H. & P.) ; John Benjamins B. V., Amsterdam ; vol. 9 ; pp. 837 - 846.
- ROUX, C. ; (1997) ; « *La valeur indiciale des fibres textiles découvertes sur un siège de voiture : Problèmes et solutions* » ; thèse de doctorat, Institut de police scientifique et de criminologie, Université de Lausanne.
- RYAN, W. J. ; BURK, K. W. ; (1972) ; « *Predictors of age in the male voice* » ; J. Acoustic. Soc. Am. ; vol. 53 ; pp. 345 (A).
- RYAN, W. J. ; BURK, K. W. ; (1974) ; « *Perceptual and acoustic correlates of aging in the speech of males* » ; Journal of Communication Disorders ; vol. 7 ; pp. 181 - 192.
- SAMBUR, M. R. ; (1975) ; « *Selection of acoustic features for speaker identification* » ; IEEE Trans. Acoust., Speech, Signal Processing ; vol. 23 ; pp. 176 - 182.
- SAMBUR, M. R. ; (1979) ; « *Speaker recognition using orthogonal linear prediction* » IN: « *Automatic Speech & Speaker Recognition* » (eds: Dixon, N. R. & Martin, T. B.) ; John Wiley & Sons, New York ; pp. 403 - 409.
- SAPIR, E. ; (1927) ; « *Speech as a personality trait* » ; Amer. J. Soc. ; vol. 32 ; pp. 892 - 895.
- SASLOVE, H. ; YARMEY, A. D. ; (1980) ; « *Long-term auditory memory: Speaker identification* » ; Journal of Applied Psychology ; no. 65 ; pp. 111 - 116.
- SAVAGE, L. J. ; (1972) ; « *The Foundations of Statistics* » ; Dover, New York.
- SAVIC, M. ; GUPTA, S. K. ; (1990) ; « *Variable parameter in speaker verification system based on hidden Markov modeling* » ; IEEE ICASSP 90 ; pp. 281 - 284.
- SCHAFFER, R. W. ; RABINER, L. R. ; (1975) ; « *Digital representations of speech signals* » ; Proc. IEEE ; vol. 63 ; pp. 662 - 677.
- SCHERER, K. R. ; (1981) ; « *Speech and emotional states* » IN: « *Speech evaluation in psychiatry* » (ed.: Darby J. K.) ; Grune & Stratton, New York ; pp. 189 - 220.
- SCHMIDT-NIELSEN, A. ; STERN, K. R. ; (1985) ; « *Identification of known voices as a function of familiarity and narrow-band coding* » ; J. Acoustic. Soc. Am. ; pp. 658 - 663.
- SCHULTZ, H. ; (1971) ; « *Der Strafrechtliche Schutz der Geheimsphäre* » ; Revue Suisse de Jurisprudence ; vol. 67 ; pp. 301 - 308.
- SCHWARTZ, M. F. ; (1968) ; « *Identification of speaker sex from isolated voiceless fricatives* » ; J. Acoustic. Soc. Am. ; vol. 43 ; pp. 1178 - 1179.
- SCHWARTZ, M. ; RINE, H. ; (1968) ; « *Identification of speaker sex from isolated whispered vowels* » ; J. Acoustic. Soc. Am. ; vol. 44 ; pp. 1736 - 1737.
- SHIPP, F. T. ; D OHERTY, T. ; HOLLIEN, H. ; (1987) ; « *Some fundamental considerations regarding voice identification* » ; J. Acoust. Soc. Am. ; vol. 82 ; no. 2 ; pp. 687 - 689.

- SHIPP, F. T. ; HOLLIEN, H. ; (1969) ; « *Perception of the aging male voice* » ; J. Speech Hearing Res. ; vol. 12 ; pp. 703 - 710.
- SHIRT, M. ; (1984) ; « *An auditory speaker recognition experiment* » ; Proceedings of the Institute of Acoustics ; vol. 6 ; no. 1 ; pp. 101 - 104.
- SIEGEL, D. M. ; (1976) ; « *Cross-examination of a 'voiceprint' expert: a blueprint for trial lawyers* » ; Crim. L. Bull. ; vol. 12 ; pp. 509 - 521.
- SILVERMAN, B. W. ; (1986) ; « *Density estimation for statistics and data analysis* » ; Chapman and Hall, London.
- SMRKOVSKI, L. ; (1976) ; « *Study of speaker identification by aural and visual identification of non-contemporary speech samples* » ; J. of the Assoc. of Official Analyt. Chem. ; vol. 59 ; pp. 927 - 931.
- SMRKOVSKI, L. ; (1997) ; « *Communication personnelle* » ; 21 janvier.
- SOLZENICYN, A. I. ; (1968) ; « *Le Premier Cercle* » ; trad. du russe par Henri-Gabriel Kybarthi ; Ex libris Lausanne.
- SOONG, F. K. ; ROSENBERG, A. E. ; (1988) ; « *On the use of instantaneous and transitional spectral information in speaker recognition* » ; IEEE Trans. ASSP ; pp. 871 - 879.
- SOONG, F. K. ; ROSENBERG, A. E. ; JUANG, B. H. ; (1987) ; « *A vector quantization approach to speaker recognition* » ; IEEE Trans. ASSP ; vol. 66 ; no. 2 ; pp. 14 - 26.
- SOONG, F. ; ROSENBERG, A. ; RABINER, L. ; JUANG, B. ; (1985) ; « *A vector quantization approach to speaker recognition* » ; ICASSP ; pp. 387 - 390.
- STEFFEN-BATOG, M. ; JASSEM, W. ; GRUSZKA-KOSCIELAK, H. ; (1970) ; « *Statistical distribution of short-term F_0 values as a personal voice characteristics* » IN: « *Speech analysis and synthesis* » (ed.: Jassem, W.) ; Polish Academy of Sciences, Warsaw ; vol. 2 ; pp. 196 - 206.
- STEINBERG, J. C. ; (1934) ; « *Application of sound measuring instruments to the study of phonetic problems* » ; J. Acoustic. Soc. Am. ; vol. VI ; pp. 16 - 24.
- STEINBERG, J. C. ; FRENCH, N. R. ; (1946) ; « *The portrayal of visible speech* » ; J. Acoust. Soc. Am. ; no. 18 ; pp. 4 - 18.
- STEVENS, K. N. ; WILLIAMS, C. E. ; CARBONELL, J. R. ; WOODS, B. ; (1968) ; « *Speaker authentication and identification: a comparison of spectrographic and auditory presentations of speech materials* » ; J. Acoustic. Soc. Am. ; vol. 44 ; pp. 1596 - 1607.
- STRATENWERTH, G. ; (1983) ; « *Schweizerisches Strafrecht, Besonderer Teil I, Straftaten gegen Individualinteressen* » ; 3^e éd., Stämpfli, Berne.
- STUDDERT-KENNEDY, M. ; (1974) ; « *The perception of speech* », IN: « *Current trends in linguistics* » (ed.: Sebeok, T. A.) ; Mouton, The Hague.
- STUDDERT-KENNEDY, M. ; (1976) ; « *Speech Perception* » IN: « *Contemporary Issues in Experimental Phonetics* » (ed.: Lass, N. J.) ; Academic Press, New York.

- SU, L. S. ; LI, K. P. ; FU, K. S. ; (1979) ; « *Identification of speakers by use of nasal coarticulation* » IN: « *Automatic Speech & Speaker Recognition* » (eds: Dixon, N. R. & Martin, T. B.) ; John Wiley & Sons, New York, USA ; pp. 378 - 384.
- SUZUKI, T. ; TANIMOTO, M. ; OSANAI, T. ; KIDO, H. ; (1994) ; « *Voice of the same male speakers twenty years apart studied on vowels* » ; 79th Annual IAI Educational Conference, Phoenix, USA.
- TARONI, F. ; AITKEN, C. G. G. ; (1996) ; « *Interpretation of Scientific Evidence* » ; Science and Justice ; vol. 36 ; no. 4 ; pp. 290 -292.
- TARONI, F. ; CHAMPOD, C. ; MARGOT, P. A. ; (1998) « *Forerunners of Bayesianism in early forensic science* » ; Jurimetrics Journal ; 38: 183-200.
- TAYLOR, H. C. ; (1933) ; « *Social agreements on personality traits as judged from speech* » ; J. Soc. Psychol. ; vol. 5 ; pp. 244 - 248.
- THEVENAZ, P. ; (1990) ; « *Reconnaissance de locuteurs indépendante du texte* » ; AGEN communications ; no. 52 ; pp. 35 - 45.
- THEVENAZ, P. ; (1993) ; « *Résidu de prédiction linéaire et reconnaissance de locuteurs indépendante du texte* » ; thèse de doctorat, Université de Neuchâtel, Suisse.
- THOMAS, K. ; (1981) ; « *Voiceprint - Myth or miracle* », IN: « *Scientific and expert evidence in criminal advocacy* » (ed.: Imwinkelried, E. J.) ; Practising Law Institute, New York City ; pp. 1015 - 1074.
- TIERNY, J. ; (1991, July 21th) ; « *Behind Monty Halls Doors, Debate and Answer?* » ; The New York Times.
- TIMOFEEV, I. N. ; SIMAKOV, V. ; (1998) ; « *Methodological basis of speaker identification within forensic phonograms investigations in the criminalistic departments of the ministry of internal affairs of Russia* » ; Proceedings of the 8th COST 250 workshop, Ankara: « *Speaker identification by man and by machine: Directions for forensic applications* » ; pp. 63 - 68.
- TIPPET, C. F. ; EMERSON, V. J. ; FEREDAY, M. J. ; LAWTON, F. ; LAMPERT, S. M. ; (1968) ; « *The evidential value of the comparison of paint flakes from sources others than vehicules* » ; J. Forensic Sci. Soc. ; vol. 8 ; pp. 61 - 65.
- TOHKURA, Y. ; (1986) ; « *A weighted cepstral distance measure for speech recognition* » ; ICASSP 86 ; pp. 761 - 764.
- TOSI, O. ; (1967) ; « *Evaluation of the voiceprint method* » ; Report to the Michigan Dept. of the State Police.
- TOSI, O. ; (1968) ; « *Speaker identification through acoustic spectrography* » ; Proc. 14th Int. Cong. on Logopedics and Phoniatrics, Paris, France.
- TOSI, O. ; (1981) ; « *Voice identification* », IN: « *Scientific and expert evidence in criminal advocacy* » (ed.: Imwinkelried, E. J.) ; Practising Law Institute New York City ; pp. 971 - 1003.
- TOSI, O. ; (1990) ; « *Historical – critical notes on voice identification / elimination* » ; J. Forensic Ident. ; vol. 40 ; no. 4 ; pp. 187 - 191.
- TOSI, O. ; NASH, E. W. ; (1973) ; « *Voiceprint identification. Rules for evidence* » ; Trial ; vol. 9 ; no. 1 ; pp. 44 - 48.
- TOSI, O. ; OYER, H. ; LASHBROOK, W. ; PEDREY, C. ; NICHOL, J. ; NASH, E. W. ; (1972A) ; « *Experiment on voice identification* » ; J. Acoust. Soc. Am. ; vol. 51 ; pp. 2030 - 2043.

- TOSI, O. ; OYER, H. ; LASHBROOK, W. ; PEDREY, C. ; NICOL, J. ; RIGGS, D. ; (1972B) ; « *Michigan state university voice identification project* » IN: « *Voice Identification Research* » ; U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice ; pp. 35 - 60.
- TSENG, B. L. ; SOONG, F. K. ; ROSENBERG, A. E. ; (1992) ; « *Continuous probabilistic acoustic map for speaker recognition* » ; ICASSP ; pp. II-161 - II-164.
- TURNER, R. F. ; RICH, V. ; ROMIG, C. H. A. ; HENNESSY, J. J. ; (1972) ; « *Some guidelines for the use of voiceprint identification technique* » IN: « *Voice Identification Research* » ; U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice ; pp. 61 - 69.
- TUTHILL, H. ; (1994) ; « *Individualization: Principles and Procedures in Criminalistics* » ; Lightning Powder Company, Inc. ; Salem, Oregon, USA.
- VAN DOMMELEN, W. A. ; (1987) ; « *The contribution of speech rhythm and pitch to speaker recognition* » ; Language and Speech ; vol. 30 ; pp. 325 - 338.
- VAN DOMMELEN, W. A. ; (1990) ; « *Acoustic parameters in human speaker recognition* » ; Language and Speech ; vol. 33 ; no. 3 ; pp. 259 - 272.
- VAN LANCKER, D. ; KREIMAN, J. ; EMMOREY, K. ; (1985A) ; « *Familiar voice recognition: patterns and parameters - part I: recognition of backwards voices* » ; Journal of Phonetics ; vol. 13 ; pp. 19 - 38.
- VAN LANCKER, D. ; KREIMAN, J. ; WICKENS, T. D. ; (1985B) ; « *Familiar voice recognition: patterns and parameters - part II: recognition of re-altered voices* » ; Journal of Phonetics ; vol. 13 ; pp. 39 - 52.
- VAN LANCKER, D. ; CUMMINGS, J. L. ; KREIMAN, J. ; DOBKINS, D. H. ; (1987) ; « *Voice discrimination and recognition are separate abilities* » ; Neuropsychologia ; vol. 25 ; pp. 829 - 834.
- VAN LANCKER, D. ; CUMMINGS, J. L. ; KREIMAN, J. ; DOBKINS, D. H. ; (1988) ; « *Phonagnosia: A dissociation between familiar and unfamiliar voices* » ; Cortex ; vol. 24 ; pp. 195 - 209.
- VAN LANCKER, D. ; KREIMAN, J. ; CUMMINGS, J. L. ; (1989) ; « *Voice perception deficits: Neuroanatomical correlates of phonagnosia* » ; Journal of Clinical and Experimental Neuropsychology ; vol. 11 ; pp. 665 - 674.
- VAN VUUREN, S. ; (1996) ; « *Comparison of text-independent speaker recognition methods on telephone speech with acoustic mismatch* » ; ICSLP, Philadelphia, PA ; no. October ; pp. 1788 - 1791.
- VIAAS (Voice Identification and Acoustic Analysis Subcommittee of the International Association for Identification) ; (1992) ; « *Voice comparison standards* » ; J. Forensic Ident. ; vol. 41 ; no. 5 ; pp. 373 - 392.
- VOIERS, W. D. ; (1964) ; « *Perceptual bases of speaker identity* » ; J. Acoustic. Soc. Am. ; vol. 36 ; no. 6 ; pp. 1065 - 1073.
- VOIERS, W. D. ; (1977A) ; « *Diagnostic evaluation of speech intelligibility* » IN: « *Speech intelligibility and speaker recognition* » (ed.: Hawley M.) ; Dowden, Hutchinson & Ross, Stroudsburg, PA, USA.
- VOIERS, W. D. ; (1977B) ; « *Diagnostic acceptability measure for speech communication systems* » ; Proc. ICASSP ; no. May ; pp. 204 - 207.
- WAGNER, I. ; (1995) ; « *A new jitter-algorithm to quantify hoarseness: an exploratory study* » ; Forensic Linguistics ; vol. 2 ; no. 1 ; pp. 18 - 27.

- WATROUS, R. L.** ; (1990) ; « *Phoneme discrimination using connectionist networks* » ; J. Acoust. Soc. Am. ; vol. 87 ; pp. 1753 - 1772.
- WELCH, E. J.** ; (1973) ; « *Voiceprint identification. A reliable index* » ; Trial ; vol. 9 ; no. 1 ; pp. 45 - 47.
- WILLIAMS, C. E.** ; (1964) ; « *The effects of selected factors on the aural identification of speakers*, IN: 'Methods for psychoacoustic evaluation of speech communication systems' » ; Dept ESD-TDR-65-153, Electronic Systems Division, Air Force Systems Command, Hanscom Field, MA.
- WILLIAMS, C. E.** ; **STEVENS, K. N.** ; **HECKER, M. H. L.** ; (1970) ; « *Acoustical manifestation of emotional speech* » ; J. Acoustic. Soc. Am. ; vol. 47 ; pp. 66.
- WOLF, J.** ; (1972) ; « *Efficient acoustic parameters for speaker recognition* » ; J. Acoust. Soc. Am. ; vol. 51 ; pp. 2044 - 2055.
- YEGNANARAYANA, B.** ; **MADHUKUMAR, A. S.** ; **RAMACHANDRAN, V. R.** ; (1992) ; « *Robust features for applications in speech and speaker recognition* » ; Proceedings of the ESCA workshop, Cannes ; pp. 97 - 101.
- YOUNG, M. A.** ; **CAMPBELL, R. A.** ; (1967) ; « *Effects of context on talker recognition* » ; J. Acoust. Soc. Am. ; no. 42 ; pp. 1250 -1254.