Thèse de doctorat de Télécom & Management SudParis dans le cadre de l'école doctorale S&I en co-accréditation avec l'Universite d'Evry-Val d'Essonne

Spécialité : Informatique

Par
M^elle HOUMANI NESMA

Thèse présentée pour l'obtention du grade de Docteur de Télécom & Management SudParis

Analyse de la qualité des signatures manuscrites en-ligne par la mesure d'entropie

Soutenue le 13 Janvier 2011 devant le jury composé de :

Pr. Nicole Vincent : Rapporteur
Pr. Laurent Heutte : Rapporteur
Pr. Hubert Cardot : Examinateur
Pr. Giuseppe pirlo : Examinateur
Dr. Zsolt Wimmer : Examinateur
Pr. Bernadette Dorizzi : Directeur de thèse
Dr. (HDR) Sonia Garcia : Co-directeur de thèse

Thèse n°2011TELE0004

Résumé

Cette thèse s'inscrit dans le contexte de la vérification d'identité par la signature manuscrite en-ligne. Notre travail concerne plus particulièrement la recherche de nouvelles mesures qui permettent de quantifier la qualité des signatures en-ligne et d'établir ainsi des critères automatiques de fiabilité des systèmes de vérification.

Nous avons proposé trois mesures de qualité faisant intervenir le concept d'entropie, très utilisé dans la théorie de l'information. Nous avons proposé une nouvelle mesure de qualité au niveau de chaque personne, appelée "Entropie personnelle", calculée sur un ensemble de signatures authentiques d'une personne. L'originalité de l'approche réside dans le fait que l'entropie de la signature est calculée en estimant les densités de probabilité localement, sur des portions, par le biais d'un Modèle de Markov Caché. Nous montrons que notre nouvelle mesure englobe les critères habituels utilisés dans la littérature pour quantifier la qualité d'une signature, à savoir : la complexité, la variabilité et la lisibilité. Par ailleurs, cette mesure permet de générer, par classification non supervisée, des catégories de personnes stables, à la fois en termes de stabilité de la signature et de complexité du tracé. En confrontant cette mesure aux performances de systèmes de vérification usuels (HMM, DTW) sur chaque catégorie de personnes générée, nous avons trouvé que les performances se dégradent de manière significative (d'un facteur 2 au minimum) entre les personnes de la catégorie "Haute Entropie" (signatures très variables et peu complexes) et celles de la catégorie "Basse Entropie" (signatures les plus stables et les plus complexes).

Nous avons ensuite proposé une nouvelle mesure de qualité basée sur l'entropie relative (appelée aussi distance de Kullback-Leibler), dénommée "Entropie Relative Personnelle" permettant de quantifier la vulnérabilité d'une personne aux attaques (bonnes imitations). Il s'agit là d'un concept original, très peu étudié dans la littérature. La vulnérabilité associée à chaque personne est calculée comme étant la distance de Kullback-Leibler entre les distributions de probabilité locales estimées sur les signatures authentiques de la personne et celles estimées sur les

imitations qui lui sont associées. Nous utilisons pour cela deux Modèles de Markov Cachés, l'un est appris sur les signatures authentiques de la personne et l'autre sur les imitations associées à cette personne. Plus la distance de Kullback-Leibler est faible, plus la personne est considérée comme vulnérable aux attaques. Cette mesure est plus appropriée à l'analyse des systèmes biométriques, du fait qu'elle englobe en plus des trois critères habituels de la littérature, la vulnérabilité aux imitations.

Enfin, nous avons aussi proposé une mesure de qualité pour les signatures imitées uniquement, ce qui est totalement nouveau dans la littérature. Cette mesure de qualité est une extension de l'Entropie Personnelle adaptée au contexte des imitations : nous avons exploité l'information statistique de la personne cible pour mesurer combien la signature imitée réalisée par un imposteur va coller à la fonction de densité de probabilité associée à la personne cible. Ainsi, nous avons défini la mesure de qualité des imitations comme étant la dissimilarité existant entre l'entropie associée à la personne à imiter et celle associée à l'imitation. Elle permet lors de l'évaluation des systèmes de vérification de quantifier la qualité des imitations, et ainsi d'apporter une information vis-à-vis de la résistance des systèmes aux attaques.

A la fin de cette thèse, nous avons aussi montré l'intérêt de notre mesure d'Entropie Personnelle pour améliorer les performances des systèmes de vérification dans des applications réelles. Nous avons montré que la mesure d'Entropie peut être utilisée pour : améliorer la procédure d'enregistrement, quantifier la dégradation de la qualité des signatures due au changement de plateforme, sélectionner les meilleures signatures de référence, identifier les signatures aberrantes, et quantifier la pertinence de certains paramètres dans le contexte de variabilité temporelle.

Abstract

This thesis is focused on the quality assessment of online signatures and its application to online signature verification systems. Our work aims at introducing new quality measures quantifying the quality of online signatures and thus establishing automatic reliability criteria for verification systems.

We proposed three quality measures involving the concept of entropy, widely used in Information Theory. We proposed a novel quality measure per person, denoted "Personal Entropy" calculated on a set of genuine signatures of such a person. The originality of the approach lies in the fact that the entropy of the genuine signature is computed locally, on portions of such a signature, based on local density estimation by a Hidden Markov Model. We show that our new measure includes the usual criteria of the literature, namely : signature complexity, signature variability and signature legibility. Moreover, this measure allows generating, by an unsupervised classification, 3 coherent writer categories in terms of signature variability and complexity. Confronting this measure to the performance of two widely used verification systems (HMM, DTW) on each Entropy-based category, we show that the performance degrade significantly (by a factor 2 at least) between persons of "high Entropy-based category", containing the most variable and the least complex signatures and those of "low Entropy-based category", containing the most stable and the most complex signatures.

We then proposed a novel quality measure based on the concept of relative entropy (also called Kullback-Leibler distance), denoted "Personal Relative Entropy" for quantifying person's vulnerability to attacks (good forgeries). This is an original concept and few studies in the literature are dedicated to this issue. This new measure computes, for a given writer, the Kullback-Leibler distance between the local probability distributions of his/her genuine signatures and those of his/-her skilled forgeries : the higher the distance, the better the writer is protected from attacks. We show that such a measure simultaneously incorporates in a single

quantity the usual criteria proposed in the literature for writer categorization, namely signature complexity, signature variability, as our Personal Entropy, but also the vulnerability criterion to skilled forgeries. This measure is more appropriate to biometric systems, because it makes a good compromise between the resulting improvement of the FAR and the corresponding degradation of FRR.

We also proposed a novel quality measure aiming at quantifying the quality of skilled forgeries, which is totally new in the literature. Such a measure is based on the extension of our former Personal Entropy measure to the framework of skilled forgeries : we exploit the statistical information of the target writer for measuring to what extent an impostor's hand-draw sticks to the target probability density function. In this framework, the quality of a skilled forgery is quantified as the dissimilarity existing between the target writer's own Personal Entropy and the entropy of the skilled forgery sample. Our experiments show that this measure allows an assessment of the quality of skilled forgeries of the main online signature databases available to the scientific community, and thus provides information about systems' resistance to attacks.

Finally, we also demonstrated the interest of using our Personal Entropy measure for improving performance of online signature verification systems in real applications. We show that Personal Entropy measure can be used to : improve the enrolment process, quantify the quality degradation of signatures due to the change of platforms, select the best reference signatures, identify the outlier signatures, and quantify the relevance of times functions parameters in the context of temporal variability.

Remerciements

Je tiens avant tout à adresser mes remerciements à ma directrice de thèse, Mme. Bernadette Dorizzi, qui m'a chaleureusement accueillie dans son laboratoire, et a accepté de diriger ma thèse. Je la remercie également pour son temps et son investissement dans tous les aspects de mon travail et de la rédaction.

Je remercie aussi Mme. Sonia Garcia qui m'a encadrée durant ces trois années de thèse. C'est grâce à nos discussions et à son expérience que ce travail a abouti à des résultats aussi intéressants. Je tiens également à la remercier pour la confiance et la sympathie qu'elle m'a témoignées tout au long de ces années de travail.

Mes remerciements vont également aux membres du Jury : Pr. Nicole Vincent, Pr. Laurent Heutte, Pr. Hubert Cardot, Pr. Giuseppe Pirlo et Dr. Zsolt Wimmer pour l'honneur qu'ils m'ont fait en participant à mon jury de thèse. Leurs remarques et suggestions lors de la lecture de mon manuscrit m'ont permis d'apporter des améliorations à la qualité de ce dernier. Je tiens à remercier plus particulièrement les deux rapporteurs de cette thèse : Pr. Nicole Vincent et Pr. Laurent Heutte, qui ont accepté de relire et évaluer ce manuscrit.

Je remercie également le Département EPH de Telecom SudParis, et en particulier Mr. Badr-Eddine Benkelfat qui m'a permis de terminer cette thèse dans de bonnes conditions de travail, et Mme. Patricia Fixot pour sa gentillesse, sa gaité et son aide lors des démarches administratives ou logistiques.

Je remercie toutes les personnes que j'ai pu rencontrer et avec lesquelles j'ai pu échanger durant mon travail de thèse : Fernando Alonso Fernandez, Jonas Richiardi, Jugurta Montalvao et tous les autres ...

Je n'oubli pas mes collègues de l'équipe Intermedia (anciens et nouveaux) : Mounim, Dijana, Emine, Anouar, Walid, Lorène, Aurélien, Sanjay, Dianle, Pierre-Olivier, Vietanh, Yosra, Guillaume, Sandra, Thierry, et tous les autres ...

Enfin, mes pensées vont à mes proches pour leur soutien constant pendant ces années. La thèse est très preneuse de temps ! Et j'avoue ne pas leur avoir consacré le

temps qu'ils méritent. Hakim ! je ne te promets pas un changement radical car l'habitude s'est installée. Je remercie plus particulièrement mes parents pour leur compréhension et surtout pour la liberté morale qu'ils m'ont toujours inculquée dans un environnement qui n'y était pas propice. Dernière pensée à *Lu. et Re.*

Table des matières

Résumé **3**

Remerciements **7**

Table des figures **15**

Liste des tableaux **23**

1 Introduction **27**
1.1 La Biométrie . . . 30
1.1.1 Propriétés des modalités biométriques . . . 30
1.1.2 Cadre général d'application de la biométrie . . . 31
1.2 Pourquoi la modalité signature ? . . . 32
1.3 Signature "en-ligne" vs. signature "hors-ligne" . . . 34
1.4 Problématiques liées à la signature en-ligne . . . 35
1.4.1 Acquisition de la signature en-ligne . . . 36
1.4.1.1 Capteurs de type stylo . . . 36
1.4.1.2 Capteurs de type tablette . . . 37
1.4.1.3 Comparaison entre les différents capteurs d'acquisition . 39
1.4.2 Variabilité de la signature en-ligne . . . 39
1.4.2.1 Variabilité temporelle . . . 39
1.4.2.2 Variabilité du style . . . 40
1.4.3 Vulnérabilité de la signature en-ligne aux attaques . . . 41
1.5 Contributions majeures de cette thèse . . . 43
1.6 Organisation de la thèse . . . 46

2 Vérification d'identité par la signature en-ligne **49**
2.1 Principe d'un système de vérification . . . 50
2.2 Mesure des performances d'un système de vérification . . . 52
2.3 Calcul de l'intervalle de confiance . . . 54
2.4 Description des bases de signatures utilisées . . . 56
2.4.1 La base Biomet . . . 56
2.4.2 La base MCYT-100 . . . 57
2.4.3 La base Philips . . . 59
2.4.4 Les bases BioSecure DS2 et DS3 . . . 60
2.4.4.1 La base BioSecure DS2 . . . 60
2.4.4.2 La base BioSecure DS3 . . . 61
2.5 Approches de classification adoptées . . . 62

2.5.1 Modèles de Markov Cachés (MMCs) . . . 64
2.5.1.1 Structure générale d'un MMC . . . 64
2.5.1.2 Les trois problèmes des MMCs . . . 66
2.5.1.3 Modélisation de la signature par un MMC . . . 67
2.5.2 Distance élastique (DTW) . . . 69
2.5.2.1 Algorithme de la distance élastique . . . 70
2.5.2.2 La distance élastique dans les systèmes de vérification . . . 72
2.5.3 Comparaison entre les deux approches MMC et DTW . . . 72
2.6 Conclusion . . . 74

3 Etat de l'art : mesures de qualité des signatures dans la littérature 75
3.1 Mesures de qualité des signatures authentiques . . . 76
3.1.1 Mesures de qualité dans le contexte en-ligne . . . 76
3.1.2 Mesures de qualité dans le contexte hors-ligne . . . 78
3.2 Mesures de qualité étendues aux "imitations" . . . 81
3.3 Catégories de personnes . . . 82
3.4 Conclusion . . . 84

4 Mesure de qualité des signatures authentiques basée sur l'entropie 85
4.1 Entropie dans la théorie de l'information . . . 86
4.1.1 Introduction à la théorie de l'information . . . 86
4.1.2 Entropie de Shannon . . . 86
4.2 L'entropie mesure de qualité en Biométrie . . . 87
4.3 Calcul de la mesure d'entropie . . . 90
4.3.1 Mesure d'entropie avec un GMM . . . 90
4.3.2 Mesure d'entropie avec un MMC . . . 93
4.4 Catégories de personnes générées par la mesure d'Entropie Personnelle . . . 96
4.5 Lien entre Entropie Personnelle, complexité et variabilité . . . 100
4.5.1 Mesure de complexité . . . 100
4.5.2 Mesure de variabilité intra-classe . . . 101
4.5.3 Entropie Personnelle vs. complexité et variabilité . . . 102
4.6 Evaluation des performances des systèmes par catégorie de personnes . . . 103
4.6.1 Description des classifieurs et protocole d'évaluation . . . 104
4.6.2 Expérimentations et résultats . . . 105
4.7 Conclusion . . . 110

5 Mesure de vulnérabilité des signatures authentiques aux imitations 113
5.1 Définition de l'entropie relative . . . 114
5.2 Mesure d'Entropie Relative avec deux MMCs . . . 115
5.3 Entropie Relative comparée à l'Entropie . . . 118
5.4 Entropie Relative et Entropie en termes de catégories de personnes . . . 119
5.4.1 Catégories de personnes d'Entropie en termes d'Entropie Relative 120
5.4.1.1 Résultats sur la base MCYT-100 . . . 120
5.4.1.2 Résultats sur la base Philips . . . 121
5.4.2 Catégories de personnes d'Entropie Relative . . . 123
5.4.2.1 Résultats sur la base MCYT-100 . . . 123
5.4.2.2 Résultats sur la base Philips . . . 126

5.4.3 Catégories de personnes générées par fusion d'Entropie et d'Entropie Relative . . . 128
5.5 Evaluation des performances par catégorie de personnes . . . 130
5.5.1 Description du classifieur et protocole d'évaluation . . . 130
5.5.2 Evaluation des performances sur la base MCYT-100 . . . 131
5.5.3 Evaluation des performances sur la base Philips . . . 134
5.6 Catégories de personnes d'Entropie et d'Entropie Relative vs. "Ménagerie Biométrique" . . . 138
5.7 Conclusion . . . 140

6 Mesure de qualité des imitations 143
6.1 Mesure de qualité des imitations . . . 144
6.1.1 Rappel sur la mesure d'Entropie Personnelle . . . 144
6.1.2 Mesure de qualité d'une imitation avec l'Entropie Personnelle . . . 145
6.2 Impact de la qualité des imitations sur les performances des systèmes de vérification . . . 146
6.2.1 Description du classifieur et protocole d'évaluation . . . 146
6.2.2 Catégorisation préliminaire des imitations . . . 147
6.2.2.1 Mesure de qualité des imitations confrontée aux imitations statiques de la base MCYT-100 . . . 147
6.2.2.2 Mesure de qualité des imitations confrontée aux différents types d'imitations de la base Philips . . . 149
6.2.2.3 Impact du protocole et des conditions d'acquisition sur la qualité des imitations dans les bases BioSecure DS2 et DS3 . . . 152
6.2.3 Catégorisation personnalisée des imitations . . . 154
6.2.3.1 Evaluation des performances par catégorie d'imitations sur les bases MCYT-100 et Philips . . . 155
6.3 Conclusion . . . 158

7 Les différentes applications de la mesure d'Entropie Personnelle 161
7.1 Campagne d'évaluation BSEC'2009 . . . 162
7.1.1 Motivation et objectifs . . . 162
7.1.2 Les participants . . . 163
7.1.3 Bases de développement et de test . . . 164
7.1.4 Description des différentes tâches d'évaluation . . . 164
7.1.4.1 Tâche 1 : impact des conditions d'acquisition sur les performances des systèmes . . . 164
7.1.4.2 Tâche 3 : impact du contenu d'information dans les signatures sur les performances des systèmes . . . 165
7.1.5 Résultats de la Tâche 1 . . . 165
7.1.6 Résultats de la Tâche 3 . . . 167
7.2 Quantification de la dégradation de qualité des signatures due au changement de plateforme . . . 169
7.3 Critère à l'enregistrement . . . 172
7.3.1 Extension des catégories de personnes avec l'Entropie Personnelle de DS2-104 à DS2-382 . . . 173
7.3.2 Evaluation des performances par catégorie de personnes de DS2-382 175

7.3.3 Proposition d'un critère à l'enregistrement . . . 176
7.4 Sélection des signatures de référence . . . 178
7.4.1 Protocole proposé pour la sélection des signatures de référence . . . 178
7.4.2 Etude des performances du système de vérification . . . 179
7.4.2.1 Description du classifieur et protocole d'évaluation . . . 180
7.4.2.2 Résultats . . . 180
7.5 Détection des signatures aberrantes . . . 182
7.5.1 Protocole proposé pour la détection des signatures aberrantes . . . 183
7.5.2 Résultats . . . 184
7.6 Etude de la robustesse des fonctions temporelles à la variabilité temporelle 187
7.6.1 Comparaison des différentes fonctions temporelles par les deux critères . . . 189
7.6.1.1 Premier critère : Entropie Personnelle . . . 189
7.6.1.2 Deuxième critère : variabilité intra-classe . . . 190
7.6.2 Protocoles d'expérimentation . . . 190
7.6.3 Analyse des résultats . . . 192
7.6.3.1 Protocole 1 vs. Protocole 2 . . . 192
7.6.3.2 Protocole 3 . . . 197
7.7 Conclusion . . . 199

8 Conclusions et perspectives 201

Bibliographie 205

A Nombre optimal de catégories de personnes 217
A.1 Procédure d'identification du nombre optimal de catégories . . . 218

B Campagne d'évaluation BSEC'2009 221
B.1 Motivation et objectifs . . . 221
B.2 Les participants . . . 222
B.3 Description des systèmes . . . 223
B.3.1 Système 1 (Escola Universitaria Politecnica de Mataro, Espagne) 223
B.3.2 Système 2 (Escola Universitaria Politecnica de Mataro, Espagne) 224
B.3.3 Système 3 (institut de recherche U1, Hongrie) . . . 225
B.3.4 Système 4 (Univ. de Seikei, Japon) . . . 226
B.3.5 Système 5 (Univ. de Ain Shams, Egypte) . . . 226
B.3.6 Système 6 (Univ. de Valladolid, Espagne) . . . 227
B.3.7 Système 7 (Univ. de Sabanci, Turquie) . . . 227
B.3.8 Systèmes 8, 9, 10, 11 et 12 (Univ. Autonoma de Madrid, Espagne) 228
B.3.8.1 Systèmes 8 et 9 . . . 228
B.3.8.2 Système 10 . . . 229
B.3.8.3 Système 11 . . . 229
B.3.8.4 Système 12 . . . 229
B.3.9 Système 13 (Univ. Wesada, Japon) . . . 229
B.3.10 Système 14 (Univ. de Magdebourg, Allemagne) . . . 230
B.3.11 Système de référence (Institut Télécom, France) . . . 231
B.4 Bases de développement et de test . . . 231
B.5 Protocoles et tâches . . . 232

B.5.1 Protocole général 232
B.5.2 Description des différentes tâches d'évaluation 232
B.5.2.1 Tâche 1 : impact des conditions d'acquisition sur les performances des systèmes 232
B.5.2.2 Tâche 2 : impact de la variabilité temporelle sur les performances des systèmes en fonction des paramètres en entrée 232
B.5.2.3 Tâche 3 : impact du contenu d'information dans les signatures sur les performances des systèmes 233
B.6 Résultats de la Tâche 1 233
B.7 Résultats de la Tâche 2 235
B.7.1 Configuration 1 : Coordonnées seulement 235
B.7.2 Configuration 2 : Coordonnées et pression 238
B.7.3 Configuration 3 : Coordonnées, pression et angles d'inclinaison . . 240
B.8 Résultats de la Tâche 3 242

C Liste des publications **245**

Table des figures

1.1 Exemple d'une signature acquise en modes "hors-ligne" (à gauche) et "en-ligne" (à droite). 34

1.2 Illustration du stylo et de la trame Anoto. 37

1.3 Les tablettes utilisées pour l'acquisition de la signature en-ligne : (a) PDA et (b) tablette graphique Wacom. 38

1.4 Exemple d'une signature (a) paraphe et d'une signature (b) cursive. . . 41

2.1 Etapes d'un système de vérification d'identité. 50

2.2 Distribution des scores des signatures authentiques et des imitations. . 52

2.3 Exemple de courbe DET. 54

2.4 Angles d'Azimut et d'Altitude utilisés pour représenter la position du stylo en 3 dimensions dans la base Biomet. 56

2.5 Exemples de signatures d'une personne appartenant à la base Biomet : (a) Signatures authentiques de la Session 1 (b) Signatures authentiques de la session 2 (c) imitations. 57

2.6 Signatures de trois personnes différentes appartenant à la base MCYT-100. 58

2.7 Les angles θx et θy utilisés pour définir la position spatiale du stylo dans la base Philips. 59

2.8 Paramètres d'un Modèle de Markov Caché. S : états, O : observations possibles, a : probabilités de transition, b : probabilités d'émission. . . 64

2.9 Modélisation de la signature par un Modèle de Markov Caché. 68

2.10 Processus général d'un système de vérification de signature en-ligne basé sur un Modèle de Markov Caché. 69

2.11 Mise en correspondance point à point entre deux signatures sans décalages temporels (à gauche), et avec décalages temporels en appliquant la distance élastique (à droite). 70

2.12 Chemin de déformation optimal. 71

2.13 Type de déplacement : insertion, substitution et suppression. 71

3.1 Exemples de signatures de MCYT-75 présentant des directions d'inclinaion a) prédominantes b) non prédominantes. 79

3.2 Graphe de lisibilité. 81

3.3 Exemple d'écritures (a) "en blocs", (b) "mixtes" et (c) "cursives". . . . 82

3.4 Catégories de personnes de la "Ménagerie Biométrique" générées en fonction des scores des données authentiques et des scores des imitations. 84

4.1 Modélisation de la signature par un Modèle à Mélange de Gaussiennes. 91

4.2 Calcul de l'entropie d'une signature avec un Modèle à Mélange de Gaussiennes. 92

4.3 Modélisation de la signature avec un Modèle de Markov Caché. 94

4.4 Calcul de l'entropie d'une signature avec un Modèle de Markov Caché. 95

4.5 Valeurs d'Entropie Personnelle pour les personnes de (a) MCYT-100 et (b) DS2-104 par catégorie de personnes : (o) Haute, (+) Moyenne et (.) Basse Entropie Personnelle. 97

4.6 Signatures de MCYT-100 à (a) Haute, (b) Moyenne et (c) Basse Entropie Personnelle. 99

4.7 Signatures de DS2-104 à (a) Haute, (b) Moyenne et (c) Basse Entropie Personnelle. 99

4.8 Exemples de signatures réparties dans la catégorie "Basse Entropie" suivant leur valeur d'Entropie calculée globalement avec un GMM. . . 100

4.9 Illustration de points (a) d'intersection simple, (b) d'intersections multiples et (c) de rebroussement. 101

4.10 Entropie Personnelle vs. complexité (à gauche) et Entropie Personnelle vs. variabilité (à droite) par catégorie de personnes de MCYT-100. . . 102

4.11 Entropie Personnelle vs. complexité (à gauche) et Entropie Personnelle vs. variabilité (à droite) par catégorie de personnes de DS2-104. 102

4.12 Courbes de performance avec les bonnes imitations (à gauche) et les imitations aléatoires (à droite) pour chaque catégorie de personnes de MCYT-100 avec le classifieur MMC. 105

4.13 Courbes de performance avec les bonnes imitations (à gauche) et les imitations aléatoires (à droite) pour chaque catégorie de personnes de MCYT-100 avec le classifieur DTW. 106

4.14 Courbes de performance avec les bonnes imitations (à gauche) et les imitations aléatoires (à droite) pour chaque catégorie de personnes de DS2-104 avec le classifieur MMC. 106

4.15 Courbes de performance avec les bonnes imitations(à gauche) et les imitations aléatoires (à droite) pour chaque catégorie de personnes de DS2-104 avec le classifieur DTW. 106

4.16 Amélioration relative $\Delta(x)$ à l'*EER* des classifieurs MMC et DTW sur MCYT-100 et DS2-104 avec les bonnes imitations, en supprimant $x\%$ de personnes contenues dans les catégories à "Haute" et "Moyenne Entropie". 109

5.1 Calcul de la divergence de Kullback-Leibler pour une personne donnée entre ses signatures authentiques et ses imitations, sur ses 10 autres signatures authentiques, après apprentissage des deux MMCs. 117

5.2 Calcul de la divergence de Kullback-Leibler pour une personne donnée entre ses signatures authentiques et ses imitations, sur ses 10 autres imitations, après apprentissage des deux MMCs. 117

5.3 Entropie Relative Personnelle en fonction de l'Entropie Personnelle pour chaque catégorie de personnes de MCYT-100 : (o) Haute, (+) Moyenne et (.) Basse Entropie Personnelle. 120

5.4 Entropie Relative Personnelle, calculée en considérant les imitations (a) *HI* et (b) *SH*, en fonction de l'Entropie Personnelle pour chaque catégorie de personnes de la base Philips : (o) Haute, (+) Moyenne et (.) Basse Entropie Personnelle. 122

5.5 Entropie Relative Personnelle en fonction de l'Entropie Personnelle pour les nouvelle catégories de personnes de la base MCYT-100 : (☆) Haute, (◇) Moyenne et (⋆) Basse Entropie Relative Personnelle. 124

5.6 Valeurs d'Entropie Personnelle des personnes de MCYT-100 appartenant uniquement à la catégorie "Basse Entropie" représentées par ◇. Parmi elles, les personnes les plus robustes aux imitations, représentant 17,91% de la catégorie "Bassse Entropie" sont désignées par ⊕. 125

5.7 Exemples de signatures de MCYT-100 (a) à "Moyenne Entropie" jugées très vulnérables aux imitations, (b) à "Basse Entropie" jugées très vulnérables aux imitations, et (c) à "Basse Entropie" jugées moyennement vulnérables aux imitations. 126

5.8 Entropie Relative Personnelle, calculée en considérant les imitations (a) *HI* et (b) *SH*, en fonction de l'Entropie Personnelle pour les nouvelles catégories de personnes de la base Philips : (☆) Haute, (◇) Moyenne et (⋆) Basse Entropie Relative Personnelle. 127

5.9 Entropie Relative Personnelle en fonction de l'Entropie Personnelle pour chaque catégorie de personnes de MCYT-100 générées par une classification hiérarchique 2D appliquées aux valeurs d'Entropie et d'Entropie Relative. 129

5.10 Courbes de performance sur les catégories de personnes de MCYT-100 générées avec (a) l'Entropie et (b) l'Entropie Relative. 131

5.11 Courbes de taux de faux rejets (a) et de taux de fausses acceptations (b) en fonction du seuil de décision sur la base MCYT-100, pour les 3 catégories générées avec la mesure d'Entropie ainsi que la catégorie contenant les personnes les plus robustes aux attaques. 133

5.12 Courbes de performance sur les catégories de personnes de la base Philips générées avec (a) l'Entropie et (b) l'Entropie Relative en considérant les imitations statiques. 135

5.13 Courbes de performance sur les catégories de personnes de la base Philips générées avec (a) l'Entropie et (b) l'Entropie Relative en considérant les imitations dynamiques. 135

5.14 Courbes de taux de faux rejets (a) et de fausses acceptations (b) en fonction du seuil de décision sur la base Philips, en considérant les imitations statiques, pour les 3 catégories générées avec la mesure d'Entropie ainsi que la catégorie contenant les personnes les plus robustes aux imitations statiques. 137

5.15 Courbes de taux de faux rejets (a) et de fausses acceptations (b) en fonction du seuil de décision sur la base Philips, en considérant les imitations dynamiques, pour les 3 catégories générées avec la mesure d'Entropie ainsi que la catégorie contenant les personnes les plus robustes aux imitations dynamiques. 137

5.16 Comparaison entre la catégorisation de personnes issue (a) de la "Ménagerie Biométrique" et (b) des deux mesures d'Entropie. 139

6.1 Exemples d'une signature authentique de la base MCYT-100 et des imitations qui lui sont associées (a) de "bonne" qualité et (b) de 'mauvaise" qualité. 148

6.2 Courbes de taux de fausses acceptations en fonction du seuil de décision sur les imitations de "bonne" et de "mauvaise" qualité de la base MCYT-100. 149

6.3 Courbes de taux de fausses acceptations en fonction du seuil de décision sur les imitations de "bonne" et de "mauvaise" qualité de la base Philips. 150

6.4 Pourcentage d'imitations de "bonne" et de "mauvaise" qualité pour chaque type d'imitations de la base Philips : statiques, dynamiques et professionnelles. 150

6.5 Courbes de taux de fausses acceptations et de faux rejets en fonction du seuil de décision sur les imitations de "bonne" et de "mauvaise" qualité de (a) DS2-120 et (b) DS3-120 contenant les mêmes personnes. 154

6.6 Courbes de taux de fausses acceptations en fonction du seuil de décision sur les imitations de "bonne", "moyenne" et "mauvaise" qualité des bases (a) MCYT-100 et (b) Philips. 156

6.7 Pourcentage d'imitations de "bonne", "moyenne" et "mauvaise" qualité pour chaque type d'imitations de la base Philips : statiques, dynamiques et professionnelles. 157

7.1 Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382 en considérant (a) les bonnes imitations et (b) les imitations aléatoires. 165

7.2 Courbes de performance des systèmes soumis à la Tâche 1 sur DS3-382 en considérant (a) les bonnes imitations et (b) les imitations aléatoires. 166

7.3 Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382, en considérant les bonnes imitations, pour les 2 catégories extrêmes d'Entropie : (a) Haute et (b) Basse Entropie. 167

7.4 Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382, en considérant les imitations aléatoires, pour les 2 catégories extrêmes d'Entropie : (a) Haute et (b) Basse Entropie. 168

7.5 Distribution des valeurs d'Entropie Personnelle sur les personnes de (a) DS2-104 et (b) DS3-104. 170

7.6 Complexité des signatures des 104 personnes lorsqu'elles sont acquises sur une plateforme fixe (DS2-104) en fonction de leur complexité quand elles sont acquises sur une plateforme mobile (DS3-104). 171

7.7 Signatures de deux personnes qui changent de catégorie sur DS2 et DS3 ; (a) de Moyenne (à gauche) à Haute Entropie (à droite), et (b) de Basse (à gauche) à Moyenne Entropie (à droite). 172

7.8 Courbes de performance sur toute la base DS2-382 et sur chaque catégorie de personnes de DS2-382 avec le classifieur MMC, en considérant (a) les bonnes imitations et (b) les imitations aléatoires. 175

7.9 Courbes de performance du classifieur DTW sur les différentes bases de référence en considérant (a) les bonnes imitations et (b) les imitations aléatoires. 181

7.10 Valeurs de la dissimilarité normalisée Q'_{sig} pour toutes les 2500 signatures authentiques disponibles dans la base MCYT-100. 184

7.11 Dendrogramme de la classification hiérarchique ascendante. 185

7.12 Exemple de signature aberrante (anomalie) détectée par notre mesure d'Entropie Personnelle. 186

7.13 Signatures aberrantes confrontées au score DTW. 187

7.14 Valeurs d'Entropie des personnes de Biomet calculées suivant les Protocoles 1 et 2 pour les combinaisons : (a) (x,y), (b) (x,y,p), (c) (x,y,Az,Alt), et (d) (x,y,p,Az,Alt). 193

7.15 Distributions de la variabilité intra-classe calculées suivant les Protocoles 1 et 2 sur la base Biomet pour les combinaisons : (a) (x,y), (b) (x,y,p), (c) (x,y,Az,Alt), et (d) (x,y,p,Az,Alt). 194

7.16 Courbes de performance sur la base Biomet avec les classifieurs DTW (à gauche) et MMC (à droite) pour chaque combinaison de paramètres, suivant le Protocole 1 (en haut) et le Protocole 2 (en bas). 195

7.17 Valeurs d'Entropie des personnes de MCYT-100 calculées suivant le Protocole 3 avec : (a) (x,y), (b) (x,y,p), (c) (x,y,Az,Alt), et (d) (x,y,p,Az,Alt). 197

A.1 Arbre hiérarchique issu de la classification des personnes de MCYT-100 suivant leur valeur d'Entropie Relative Personnelle. A droite, la projection des niveaux de nœuds. 219

A.2 Indices de validité calculés pour chaque valeur de nombre de catégories k : (a) C-index, (b) Weighted intra-inter index, (c) Krzanowski-Lai index, et (d) RMSSTD Group. 220

B.1 Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382 en considérant (a) les bonnes imitations et (b) les imitations aléatoires. 234

B.2 Courbes de performance des systèmes soumis à la Tâche 1 sur DS3-382 en considérant (a) les bonnes imitations et (b) les imitations aléatoires. 234

B.3 Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires, sur la Session 2 de la base DS2-382 en considérant uniquement les coordonnées en entrée des systèmes. 236

B.4 Courbes de performance des systèmes avec (a) les bonnes imitations et (b) les imitations aléatoires, sur la Session 1 de la base DS3-382. 237

B.5 Courbes de performance des systèmes avec (a) les bonnes imitations et (b) les imitations aléatoires, sur la Session 2 de la base DS3-382. 237

B.6 Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires sur la Session 1 de DS2-382, en considérant les coordonnées et la pression en entrée des systèmes. 239

B.7 Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires sur la Session 2 de DS2-382, en considérant les coordonnées et la pression en entrée des systèmes. 239

B.8 Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires sur la Session 1 de DS2-382, en considérant les coordonnées, la pression et les angles d'inclinaison en entrée des systèmes. 241

B.9 Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires sur la Session 2 de DS2-382, en considérant les coordonnées, la pression et les angles d'inclinaison en entrée des systèmes. 241

B.10 ourbes de performance des systèmes soumis à la Tâche 1 sur DS2-382, en considérant les bonnes imitations, pour les 2 catégories extrêmes d'Entropie : (a) Haute et (b) Basse Entropie. 242

B.11 Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382, en considérant les imitations aléatoires, pour les 2 catégories extrêmes d'Entropie : (a) Haute et (b) Basse Entropie. 243

Liste des tableaux

4.1 Distribution des personnes de MCYT-100 en chaque catégorie d'entropie 97

4.2 Distribution des personnes de MCYT-100 dans chaque catégorie d'Entropie Personnelle. 98

4.3 Distribution des personnes de DS2-104 dans chaque catégorie d'Entropie Personnelle. 98

4.4 Taux d'erreur à l'*EER* et intervalles de confiance par catégorie de personnes de MCYT-100 avec le classifieur MMC. 107

4.5 Taux d'erreur à l'*EER* et intervalles de confiance par catégorie de personnes de MCYT-100 avec le classifieur DTW. 107

4.6 Taux d'erreur à l'*EER* et intervalles de confiance par catégorie de personnes de DS2-104 avec le classifieur MMC. 108

4.7 Taux d'erreur à l'*EER* et intervalles de confiance par catégorie de personnes sur DS2-104 avec le classifieur DTW. 108

4.8 Amélioration relative à l'*EER* des classifieurs MMC et DTW sur MCYT-100 et DS2-104, en supprimant toutes les personnes des catégories "Haute" et "Moyenne Entropie". 110

5.1 Matrice de corrélation entre l'Entropie et l'Entropie Relative sur la base MCYT-100. 121

5.2 Distribution de personnes de MCYT-100 dans chaque catégorie d'Entropie Personnelle et d'Entropie Relative Personnelle. 124

5.3 Distribution des personnes de la base Philips dans chaque catégorie d'Entropie et d'Entropie Relative avec les deux types d'imitations (statiques et dynamiques). 127

5.4 Distribution des personnes de MCYT-100 dans chaque catégorie générée par fusion de l'Entropie et l'Entropie Relative. 129

5.5 Taux d'erreur à l'*EER* et intervalles de confiance pour chaque catégorie d'Entropie et d'Entropie Relative de la base MCYT-100. 132

5.6 Taux d'erreur à l'*EER* et intervalles de confiance pour chaque catégorie d'Entropie et d'Entropie Relative de la base Philips en considérant uniquement les imitations statiques. 135

5.7 Taux d'erreur à l'*EER* et intervalles de confiance pour chaque catégorie d'Entropie et d'Entropie Relative de la base Philips en considérant uniquement les imitations dynamiques. 136

6.1 Distribution des imitations statiques de MCYT-100 en imitations de "bonne" et "mauvaise" qualité. 148

6.2 Distribution des imitations de la base Philips en imitations de "bonne" et de "mauvaise" qualité. 149

6.3 Distribution des imitations de la base DS2-120 en imitations de "bonne" ou "mauvaise" qualité. 153

6.4 Distribution des imitations de la base DS3-120 en imitations de "bonne" ou "mauvaise" qualité. 153

6.5 Distribution des imitations des bases MCYT-100 et Philips en "bonne", "moyenne" et "mauvaise" qualité. 155

7.1 Liste des équipes participant à la campagne d'évaluation BSEC'2009. . 163

7.2 Taux d'erreur à l'*EER* des systèmes soumis à la Tâche 1 sur la Session 1 de DS2-382 et DS3-382 avec les deux types d'imitations. 166

7.3 Taux d'erreur à l'*EER* sur la Session 1 de DS2-382 pour les deux catégories extrêmes de personnes avec les deux types d'imitations. . . . 168

7.4 Pourcentage des personnes de DS2-104 et DS3-104 dans chaque catégorie d'Entropie Personnelle. 170

7.5 Distribution des personnes de DS2-382 dans chaque catégorie d'Entropie générée suivant les "Prototypes d'Entropie" calculés au préalable sur DS2-104. 174

7.6 Taux d'erreur à l'*EER* et intervalles de confiance pour chaque catégorie de personnes de DS2-382 avec le classifieur MMC en considérant les deux types d'imitations (bonnes et aléatoires). 175

7.7 Taux d'erreur à l'*EER* du classifieur DTW sur chaque base de références en considérant les deux types d'imitations (bonnes et aléatoires). . . . 181

7.8 Taux d'erreur à l'*EER* et intervalles de confiance à 95% pour chaque combinaison de paramètres sur la base Biomet avec les classifieurs DTW et MMC suivant le Protocole 1. 196

7.9 Taux d'erreur à l'*EER* et intervalles de confiance à 95% pour chaque combinaison de paramètres sur la base Biomet avec les classifieurs DTW et MMC suivant le Protocole 2. 196

7.10 Taux d'erreur à l'*EER* pour chaque combinaison de paramètres sur MCYT-100 avec les classifieurs DTW et MMC suivant le Protocole 3. . 198

B.1 Liste des équipes participant à la campagne d'évaluation BSEC'2009. . 223

B.2 Taux d'erreur à l'*EER* sur la Session 1 de DS2-382 et DS3-382 avec les deux types d'imitations, en considérant les coordonnées en entrée des systèmes. 235

B.3 Taux d'erreur à l'*EER* sur la Session 2 de DS2-382 avec les deux types d'imitations, en considérant les coordonnées en entrée des systèmes. . . 236

B.4 Taux d'erreur à l'*EER* sur les deux Sessions de DS3-382 avec les deux types d'imitations. 238

B.5 Taux d'erreur à l'*EER* sur les deux Sessions de DS2-382 avec les deux types d'imitations, en considérant les coordonnées et la pression en entrée des systèmes. 240

B.6 Taux d'erreur à l'*EER* sur les deux Sessions de DS2-382 avec les deux types d'imitations, en considérant les coordonnées, la pression et les angles d'inclinaison en entrée des systèmes. 242

B.7 Taux d'erreur à l'*EER* sur la Session 1 de DS2-382 pour les deux catégories extrêmes de personnes avec les deux types d'imitations. . . . 243

Chapitre 1

Introduction

Cette thèse s'inscrit dans le domaine de la biométrie, et plus particulièrement dans le contexte de la vérification d'identité par la signature manuscrite en-ligne.

La signature manuscrite est une modalité comportementale qui repose sur un geste naturel et des mouvements instinctifs lors de sa réalisation. Son utilisation répandue depuis plusieurs siècles fait d'elle le moyen le plus accepté socialement et juridiquement pour la vérification d'identité. Elle est utilisée communément sur des chèques, des lettres, des contrats, toutes sortes de documents administratifs pour s'identifier ou manifester la bonne foi et la propre volonté de la personne.

Dû au grand nombre de documents (chèques bancaires par exemple) qui doivent être authentifiés dans un temps limité, la vérification manuelle des signatures manuscrites est très peu réaliste. Cela a mené à la recherche de systèmes de vérification de signatures automatiques.

Or, la vérification de l'identité par la signature manuscrite se trouve confrontée à deux problématiques majeures : la première réside dans le fait que la signature, comme toute modalité comportementale, présente une forte variabilité intra-classe (la signature varie d'un instant à l'autre pour une même personne), ainsi qu'une variabilité au cours du temps ; la deuxième problématique concerne sa vulnérabilité aux attaques. En effet, notre signature n'est malheureusement pas secrète et un faussaire potentiel peut facilement reproduire l'image de la signature. De ce fait, la vérification automatique par la signature manuscrite demande la mise en place de systèmes dont l'efficacité dépend de leur capacité à différencier les signatures authentiques des imitations.

En pratique, il existe deux modes de réalisation de signature : le mode "hors-ligne" et le mode "en-ligne". Le premier n'utilise que l'image statique de la signature et donc son information spatiale. Le deuxième utilise à la fois l'information spatiale et dynamique, décrivant ainsi le geste pratiqué lors de la réalisation de la signature grâce à une acquisition sur un dispositif (tablettes, smartphones,...). La signature "en-ligne" est donc plus riche en information, plus spécifique à la personne, moins variable et plus difficile à falsifier. Pour toutes ces raisons, la signature en-ligne a fait l'objet de notre étude.

Cette thèse s'insère dans la continuité des travaux effectués au sein de l'équipe Intermedia du département EPH de Télécom SudParis sur la vérification de signatures en-ligne. En fait, plusieurs types de systèmes de vérification d'identité automatiques par la signature en-ligne ont été développés et évalués sur différentes bases de données au sein de cette équipe. De très bons résultats ont été obtenus, mais nous avons remarqué que les performances des systèmes sont très dépendantes des bases utilisées pour un même classifieur. De surcroît, selon les bases utilisées, les résultats de diverses méthodes de classification étaient différents voire contradictoires. Ceci est dû aux différences qui existent entre les bases de signatures en terme de nombre et type de personnes, de protocole et d'environnement d'acquisition.

D'où l'intérêt de développer un nouvel axe de recherche consacré à l'investigation de mesures qui permettent de *"quantifier la qualité"* des signatures en-ligne, et d'établir ainsi des critères de fiabilité des systèmes de vérification en fonction des différents facteurs qui influent sur la qualité des signatures : qualité de l'environnement d'acquisition, risque d'attaques de haut niveau (imitations de bonne qualité), etc. C'est le problème que nous nous posons dans cette thèse.

Au tout début de notre travail de thèse, nous avons fait un état des lieux des travaux dans le domaine de la biométrie qui abordent cette problématique. Nous avons constaté que les mesures de qualité sont couramment utilisées dans les modalités physiologiques, tel que le visage où la qualité est définie comme étant la netteté, le niveau du contraste et l'équilibre de l'illumination [84] ; pourtant, elles restent peu utilisées et mal définies dans les modalités comportementales, telle que la signature en-ligne. Certains travaux menés dans le cadre de la signature manuscrite ont lié les performances des systèmes de vérification à la complexité du tracé de la signature, à sa variabilité ou à sa lisibilité. De plus, la qualité des signatures est souvent mesurée de façon subjective, de ce fait, elle ne pourrait

pas être a priori utilisée comme une information supplémentaire dans un système automatique bien qu'elle soit pourtant bien utile à l'évaluation du système.

Notre but étant de mesurer la qualité des signatures en-ligne, il était alors nécessaire avant toute chose de déterminer ce qu'est la qualité d'une signature ? Nous sommes parti du principe que la qualité d'une signature en-ligne peut être définie comme étant le contenu d'information dans cette signature. Ceci nous a amené à nous intéresser de plus près au domaine de la théorie de l'information.

Ce que nous proposons dans cette thèse est d'introduire de nouvelles mesures de qualité basées sur le concept d'entropie issu de la théorie de l'information. Nous avons proposé un calcul original de l'entropie d'une signature dynamique en estimant localement, sur des portions de la signature, les densités de probabilité en faisant appel aux Modèles de Markov Cachés. Nos études ont montré que cette mesure est très bien adaptée au contexte de la signature manuscrite en-ligne et qu'elle permet de quantifier une signature par un seul critère qui englobe d'autres critères intuitifs (complexité, variabilité, lisibilité).

Comme l'évaluation des systèmes de vérification de signatures est effectuée sur des bases contenant des signatures authentiques et des imitations, nous avons utilisé ce concept d'entropie pour définir trois types de mesures, à savoir : mesure de qualité des signatures authentiques (Chapitre 4), mesure de leur vulnérabilité aux imitations (Chapitre 5), et enfin mesure de qualité des imitations (Chapitre 6).

Après avoir défini de manière générale le contexte et les motivations de cette thèse, puis donné brièvement la ligne directrice de notre travail, dans la suite de ce premier chapitre, nous aborderons plus en détail certains concepts. Nous commencerons tout d'abord par définir ce qu'on entend par biométrie et son intérêt applicatif potentiel dans nos sociétés. Par la suite, nous présenterons la signature manuscrite et plus particulièrement le contexte en-ligne. Nous nous intéresserons ensuite aux problématiques et contraintes liées à l'utilisation de la signature en-ligne comme moyen de vérification de l'identité d'un individu. Ceci permettra de définir le contexte et les motivations de notre travail, et de situer notre action dans le cadre de l'étude des mesures de qualité des signatures. La dernière partie de ce chapitre s'attachera à déterminer comment notre travail contribue à atteindre nos objectifs en décrivant la démarche adoptée pour chacune des trois parties autour desquelles cette thèse est articulée.

1.1 La Biométrie

De nos jours, déterminer de manière à la fois efficace et exacte l'identité d'un individu est devenu un problème critique dans notre société. En effet, devant la croissance exponentielle des communications, tant en volume qu'en diversité (passages de frontières, transactions financières, accès aux services...), et les risques associés en terme d'usurpation d'identité, il est devenu nécessaire de contrôler et de vérifier l'identité des acteurs de ces échanges. D'autant plus que l'importance des enjeux motive les fraudeurs à mettre en échec les systèmes de sécurité existants.

Plusieurs alternatives ont été envisagées pour vérifier l'identité des personnes en utilisant différents types d'informations : ce que l'on possède comme la "clef" et le "badge" qui peuvent être perdus ou utilisés par des tiers non autorisés ; ce qu'on sait comme le "code PIN" ou le "mot de passe" qui peuvent être oubliés ou facilement déchiffrables. Devant ces limites, la biométrie est apparue comme la solution à envisager afin d'éviter le détournement ou le vol d'informations sensibles.

Le terme "biométrie" désigne dans un sens très large l'étude quantitative des êtres vivants, mais dans notre contexte, la biométrie désigne plus spécifiquement l'ensemble des techniques d'identification ou d'authentification des individus par des caractéristiques qui leur sont propres. Ces caractéristiques peuvent être biologiques telles que l'odeur, le sang, la salive, l'urine et l'ADN ; morphologiques telles que les empreintes digitales, l'iris, le visage et la main ; et même comportementales telles que la signature manuscrite, la voix et la démarche.

1.1.1 Propriétés des modalités biométriques

Ces caractéristiques dites biométriques ont un caractère personnel et ne peuvent pas être facilement volées, falsifiées, ou partagées. Ainsi, elles sont plus fiables et sécurisées pour la reconnaissance de personnes que les méthodes traditionnelles basées sur la connaissance ou la possession. Chaque caractéristique, physiologique ou comportementale, peut être utilisée comme une modalité biométrique dès lors qu'elle possède les propriétés suivantes :

- l'*universalité*, signifie que toutes les personnes de la population à identifier devraient posséder cette caractéristique biométrique ;

- l'*unicité*, indique que deux personnes ne peuvent posséder exactement la même caractéristique ;

- la *permanence*, signifie que la caractéristique doit subsister durant la vie d'un individu afin de permettre une vérification au cours du temps.

- la *collectabilité*, se rapporte au fait que la caractéristique doit être facilement mesurable, afin de permettre l'enregistrement de la donnée biométrique et les comparaisons futures.

- l'*acceptabilité*, signifie que les individus doivent être prêts à enregistrer cette donnée biométrique.

- la *non-reproductibilité*, porte sur la facilité ou non à falsifier et reproduire la donnée biométrique.

Chaque modalité biométrique a ses forces et ses faiblesses, et possède effectivement ces propriétés avec des degrés différents. Le choix de l'utilisation d'une modalité biométrique dépend généralement des besoins de l'application à traiter.

1.1.2 Cadre général d'application de la biométrie

La biométrie n'est pas vraiment récente, ses applications remontent à la fin du 19ème siècle avec l'utilisation des empreintes digitales pour l'identification des délinquants par la police. Aujourd'hui, le champ d'application de la biométrie couvre potentiellement tous les domaines de sécurité où il est nécessaire de s'assurer de l'identité des personnes, et plus particulièrement au passage des frontières. En effet, la mise en place du passeport biométrique en 2009 constitue une étape importante pour les gouvernements dans leur volonté de contrôler les flux migratoires.

Malgré l'engouement dont jouit la biométrie à l'heure actuelle, favorisé par un phénomène de mode diffusé surtout par les séries policières, son usage reste relativement restreint dans notre quotidien, surtout en Europe. On a encore le plus souvent recours à des méthodes cryptographiques basées sur la mémorisation de codes, comme c'est le cas pour les cartes bancaires. Cela est justifié par le fait que l'identification d'une personne par la biométrie repose sur une mesure de similarité entre deux données acquises à des moments différents (enregistrement et test) et ces données ne sont jamais tout à fait semblables même si elles proviennent de la même personne. En effet, les caractéristiques biométriques, basées sur la physiologie ou le comportement des individus, évoluent et changent au cours du temps. De ce fait, il est possible de faire des erreurs avec la biométrie, ce qui n'est pas possible avec

un code ou un mot de passe. De plus, l'utilisation de caractéristiques propres à la personne pour s'identifier n'est pas encore véritablement entrée dans les mœurs en dehors de son utilisation policière. Par ailleurs, les problèmes relatifs à la protection de la vie privée et des données personnelles sont un frein important à l'application des techniques biométriques à grande échelle. Le coût important des technologies biométriques a aussi longtemps limité leur développement.

1.2 Pourquoi la modalité signature ?

Les performances des systèmes biométriques qui reposent sur des modalités différentes ne sont pas équivalentes ; on constate en fait que certaines modalités morphologiques, comme l'iris et l'empreinte digitale, sont plus fiables que les modalités comportementales, comme la voix et la signature manuscrite.

En dépit de ses faibles performances, la signature manuscrite est la modalité comportementale la plus aboutie, et ses avantages sont nombreux.

En comparaison avec les modalités physiologiques, de nature morphologique ou biologique, l'utilisation de la signature manuscrite comme moyen de vérification d'identité constitue un bon compromis entre le niveau de sécurité (fiabilité), la facilité d'utilisation et le prix (la signature ne nécessite pas un coût supplémentaire élevé pour le capteur), sans parler de sa forte acceptabilité dans la population.

A propos du bien-fondé de son utilisation, on peut ajouter qu'en apposant sa signature manuscrite, chaque personne exprime -au sens propre du terme- l'empreinte de sa personnalité. D'ailleurs l'observation de personnes en train de signer conduit à émettre l'hypothèse que le geste pratiqué lors de la réalisation de la signature est plus un mouvement instinctif qu'un acte conscient. Cette conjecture implique que certaines caractéristiques de la signature sont stables et donc constantes pour une personne donnée.

En ce qui concerne sa fiabilité, chaque signature est unique ! En effet, elle reflète des mécanismes propres à la personne, de nature autant physiologique que biomécanique, qui sont affinés avec le temps du fait d'une utilisation régulière. Toutefois, deux signatures d'une même personne ne peuvent jamais être exactement identiques sauf s'il s'agit d'une copie.

Aussi, signer est un geste de la vie courante et de ce fait n'est pas vécu comme une contrainte invasive pour l'utilisateur. A ce jour, l'une des méthodes les

plus fréquemment acceptées et utilisées pour permettre la non répudiation ou la preuve d'engagement de l'individu est sa signature manuscrite. Son utilisation pour l'authentification est autant habituelle qu'acceptée, contrairement à certaines modalités dont l'acquisition peut être vécue comme invasive (iris, empreinte). Ainsi, l'adhésion des utilisateurs à ce moyen de vérification d'identité est acquise.

De surcroît, la signature propre de la personne n'est jamais apposée involontairement. Dans le cas où la personne est contrainte à signer, on verra sa signature changer à cause du stress et la nervosité engendrée par la situation. Aussi, si on veut utiliser frauduleusement le système de vérification, on ne peut qu'imiter la signature mais pas reproduire exactement la même signature, comme c'est le cas pour les empreintes digitales qui peuvent être reproduites sur un faux doigt.

Par ailleurs, l'acte de signer n'a pas de connotation particulière, contrairement à l'empreinte digitale qui a une connotation d'investigation criminelle. Finalement, un dernier avantage de cette modalité, du point de vue de l'utilisateur, est que ce dernier peut s'il le désire changer sa signature.

Malgré les avantages associés à la signature manuscrite, cette dernière reste peu utilisée dans des applications biométriques. Comme toute modalité comportementale, la signature manuscrite présente certains inconvénients. Ses principaux inconvénients sont liés à la grande variabilité, à savoir une variabilité intra-classe (pour un même individu) et une variabilité au cours du temps. Cette variabilité peut être accentuée par des changements émotionnels ou environnementaux chez l'utilisateur. Le stress, la fatigue ou un environnement perturbé, par exemple, peuvent affecter le comportement de l'individu et ainsi perturber le résultat du test de reconnaissance. De plus, comme la signature est un moyen d'authentification très répandu, des traces de signatures sont facilement disponibles, elle est donc sujette aux imitations et il est possible de produire de très bonnes imitations après plusieurs entraînements.

La modalité signature a été retenue pour notre étude pour diverses raisons. Plusieurs systèmes automatiques de vérification d'identité par signature manuscrite ont déjà été développés par notre équipe dans le passé. Nous disposons aussi de plusieurs bases de signatures avec différents protocoles d'acquisition. Il aurait été alors dommage de ne pas profiter de ces acquis, et de ne pas pousser le travail scientifique sur la signature manuscrite en vue d'introduire des critères quantitatifs de fiabilité de ces systèmes de vérification sur ces différentes bases de signatures. De plus, la signature est une modalité purement comportementale qui présente une forte variabilité pour une même personne. De ce fait, elle est souvent sujette

à de fortes altérations. Ainsi, cela a un sens d'étudier sa qualité et de faire le lien entre la qualité d'une signature et les performances que l'on peut attendre d'un système de vérification d'identité sur cette signature.

Comme déjà mentionné, suivant la méthode de capture de la signature, on distingue deux catégories de signatures manuscrites : "hors-ligne" et "en-ligne". La section suivante décrit les différences existant entre ces deux catégories.

1.3 Signature "en-ligne" vs. signature "hors-ligne"

Dans un système dit "hors-ligne", la signature est d'abord réalisée sur un support papier puis numérisée de façon différée à l'aide d'un scanner ou d'une caméra numérique. D'où le terme "hors-ligne". La signature est alors assimilée à une image en niveaux de gris (voir Figure 1.1.a). Ainsi, seulement la forme de la signature est disponible, et seules les données statiques décrivant la géométrie de la signature sont prises en compte. C'est le cas notamment pour les systèmes de vérification de chèques, de contrats ou de formulaires administratifs.

En mode "hors-ligne", on ne dispose pas de paramètres représentant la dynamique de la signature. Afin de rendre les systèmes d'authentification plus fiables, cette dernière peut être générée de façon indirecte par le biais de certaines informations. Par exemple, l'épaisseur du trait ou la variation d'intensité du niveau de gris dans la signature décrivent les différentes coulées d'encre sur le papier, et peuvent donc être des indicateurs de la pression exercée par le signataire sur le papier.

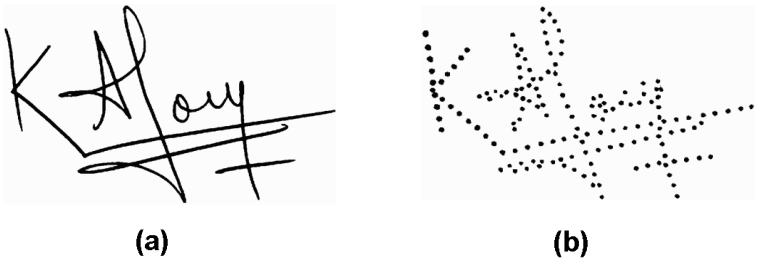

FIGURE 1.1: Exemple d'une signature acquise en modes "hors-ligne" (à gauche) et "en-ligne" (à droite).

Cependant, la vérification de l'identité par la signature "hors-ligne" a montré depuis longtemps ses limites, car reproduire l'image de la signature reste un processus relativement simple, et donc un système de vérification automatique a du mal à discriminer les signatures imitées des signatures authentiques. La signature

en tant qu'image devenant vite insuffisante pour permettre une authentification fiable, il était nécessaire d'extraire plus d'informations propres à la personne qui signe. D'où le recours à la signature "en-ligne" qui se base sur l'étude du mouvement, ou autrement dit sur le "geste" pratiqué lors de la réalisation de la signature.

Contrairement à la signature "hors-ligne", la signature "en-ligne" est numérisée directement par un dispositif qui permet de l'échantillonner à intervalles de temps réguliers, pendant que la personne signe (voir Figure 1.1.b). Le signal ainsi récupéré est une séquence temporelle.

En fonction du type de capteur, soit la trajectoire de la signature est seule échantillonnée et sauvegardée, sinon, d'autres informations sont aussi acquises, comme la pression et les angles d'inclinaison du stylo. Les signaux échantillonnés permettent ainsi de décrire le geste du signataire. De ce fait, la signature "en-ligne" prend en compte non seulement l'aspect spatial de la signature (la forme de la signature) mais aussi l'aspect dynamique associé au geste pratiqué lors de la réalisation de la signature. Par conséquent, le mode "en-ligne" est plus riche en information que le mode "hors-ligne", plus propre à la personne, et donc plus discriminant.

L'avantage indéniable de la signature en-ligne est d'une part l'impossibilité pour un imposteur, qui ne voit que la forme de la signature, de reproduire les informations dynamiques de cette signature. D'autre part, toutes les informations dynamiques se retrouvent d'une manière plus ou moins stable d'une signature à l'autre chez une même personne, et peuvent être donc utilisées pour rendre le processus de vérification plus fiable. C'est pourquoi dans cette thèse nous avons fait le choix de travailler sur la signature "en-ligne".

1.4 Problématiques liées à la signature en-ligne

Bien que la signature "en-ligne" soit plus fiable que la signature "hors-ligne", elle reste tout de même exposée à certaines problématiques qui peuvent fortement affecter les performances des systèmes de vérification.

Nous verrons dans la suite que certains problèmes sont liés au capteur et aux conditions d'acquisition pouvant altérer la qualité du signal temporel associé à la signature en-ligne. D'autres, sont liés à l'utilisateur lui-même : comme la signature est purement comportementale, elle présente une variabilité intra-classe

et une variabilité au cours du temps. D'autre part, nous aborderons le problème des imitations car bien que la signature "en-ligne" soit plus difficile à falsifier que la signature "hors-ligne", elle reste exposée à différents types d'attaques, dont certaines sont très évoluées.

1.4.1 Acquisition de la signature en-ligne

L'acquisition de la signature est une phase très importante car la qualité de la signature dépend fortement du capteur d'acquisition (sa résolution, son temps de réponse, sa facilité d'utilisation, etc.). De ce fait, les dispositifs d'acquisition de la signature doivent offrir une ergonomie suffisante pour que les usagers les utilisent sans fournir d'effort supplémentaire. De plus, comme beaucoup de personnes, habituées à signer sur papier, adaptent ou modifient leur manière de signer lors de l'acquisition, un temps d'adaptation au support numérique est donc nécessaire afin d'obtenir une stabilité suffisante de la signature.

Plusieurs types de capteurs d'acquisition de la signature en-ligne existent. La différence entre eux réside dans la technologie qu'ils emploient. Nous distinguons deux catégories de capteurs : capteurs de type stylo, et d'autres de type tablette.

1.4.1.1 Capteurs de type stylo

Dans cette catégorie, le capteur le plus répandu est le stylo Anoto [56, 27, 103, 41] mis au point par la société suédoise *Anoto* [56]. Ce stylo permet le transfert de textes manuscrits sur papier vers un ordinateur ou un téléphone portable.

Son concept repose sur l'utilisation couplée d'un stylo bille muni d'une caméra digitale intégrée et d'un papier spécifique breveté (*Pattern Anoto*) sur lequel a été imprimée une trame imperceptible à l'œil nu (voir Figure 1.2).

L'originalité de cette trame est qu'elle repose sur un algorithme mathématique très élaboré. Elle est constituée d'une immense grille carrée, composée de minuscules points espacés de $0,3mm$. Ces points sont, à chaque fois, légèrement décalés des intersections de la grille, soit un peu vers le haut, le bas, la gauche ou la droite (voir Figure 1.2). Ensemble, ils forment de minuscules motifs géométriques qui servent de repères pour le positionnement du stylo [56].

La surface minimale, pour que le stylo puisse se repérer, est un carré de 6 points de côté, soit environ une surface de $3,24mm^2$. Les différentes combinaisons

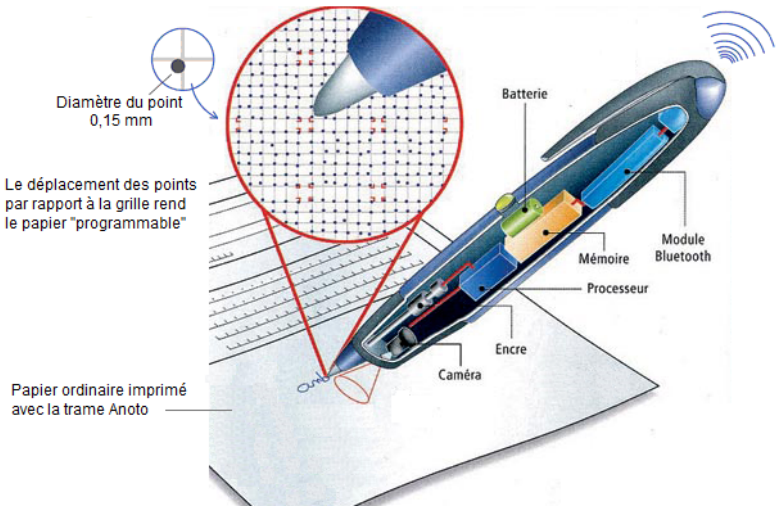

FIGURE 1.2: Illustration du stylo et de la trame Anoto.

de motifs permettent de déterminer chacune des positions du stylo de façon unique et ce, sur une surface équivalente à près de 60 millions de km^2 [57], soit plus de la surface de l'Europe et de l'Asie réunies !

1.4.1.2 Capteurs de type tablette

Dans cette catégorie, deux types de tablettes sont utilisés à nos jours : les tablettes passives et les tablettes actives [103, 41].

Tablettes passives Les tablettes dites "passives" [103, 41], telles que les PDAs (*Personal Digital Assistants*) (voir Figure 1.3.a), sont souvent associées à un stylet passif, et il n'existe aucune communication entre ce dernier et la tablette. Ces tablettes sont basées sur la technologie tactile [55, 103, 41] qui peut être essentiellement capacitive ou résistive [55, 41] suivant l'usage que l'on veut en faire.

La technologie capacitive est basée sur un changement de capacité électrique à l'endroit du contact [55, 41]. Les écrans tactiles basés sur cette technologie sont adaptés à un usage public. Ils ont l'avantage d'avoir une longue durée de vie, un temps de réponse au contact très rapide, et une transparence élevée de l'affichage. Cependant, ils ne réagissent qu'au contact d'un doigt nu. Un doigt ganté comme un stylet reste inopérant. Ces écrans peuvent donc être nettoyés sans nécessiter une interruption du système et sans que cela ne provoque l'entrée de données erronées.

La technologie résistive quant à elle est basée sur la création d'une tension lors d'un contact de deux couches conductrices [55, 41]. Les écrans tactiles basés

sur cette technologie sont adaptés à un usage privé. Ils sont très économiques et sensibles à toute forme de toucher avec le doigt ou un stylet. Néanmoins, ils exigent d'épaisses couches conductrices, ce qui diminue la transparence et la qualité d'affichage. De plus, la surface de l'écran se raie facilement.

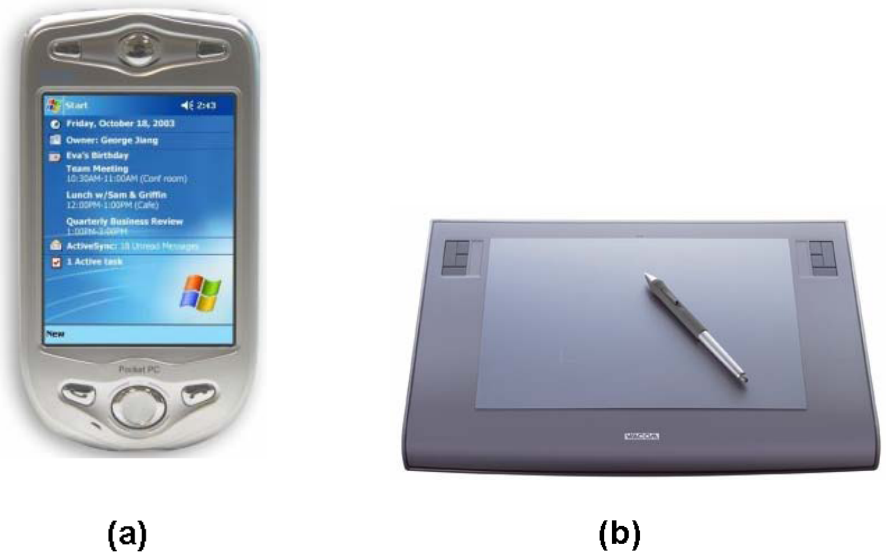

FIGURE 1.3: Les tablettes utilisées pour l'acquisition de la signature en-ligne : (a) PDA et (b) tablette graphique Wacom.

Tablettes actives Contrairement aux PDAs où le stylet, s'il est utilisé, n'est qu'un outil passif, les tablettes dites "actives" nécessitent l'utilisation d'un stylet électronique (stylo) qui communique en permanence avec une tablette graphique (voir Figure 1.3.b). Par conséquent, les tablettes actives ne détectent ni la présence d'un stylet passif, ni l'effleurement du doigt [59, 103, 41].

Les tablettes actives les plus répandues sont les tablettes graphiques développées par la société japonaise *Wacom* [59, 103, 41]. Ces tablettes de haute gamme utilisent la technologie électromagnétique et se basent sur le transfert d'énergie sans fil. En effet, équipé d'un circuit résonant, le stylet électronique n'a besoin pour son fonctionnement ni d'un cordon obstructif ni d'une batterie, car il est alimenté par la tablette via la technologie de résonance électromagnétique [58, 103, 41]. De ce fait, le stylet a l'avantage d'être mince et léger, et offre une grande liberté de mouvement. De plus, grâce à la technologie de résonance électromagnétique, la tablette graphique a la capacité d'enregistrer la position du stylo sans que ce dernier touche la surface de tablette.

1.4.1.3 Comparaison entre les différents capteurs d'acquisition

La tablette graphique est le capteur le plus apprécié par les professionnels car elle permet l'acquisition de 5 fonctions temporelles : les coordonnées du stylo sur le repère associé à la tablette, la pression et les angles d'inclinaison du stylo. Aussi, elle offre une grande résolution et une meilleure précision.

Le PDA et le stylo Anoto sont les capteurs qui offrent le moins de possibilités du fait que les données acquises se limitent uniquement à la position du stylo et le temps. De plus, leur puissance de calcul ainsi que la résolution sont relativement faibles, ce qui induit, par exemple dans le cas du PDA, une perte d'information dans le tracé de la signature. Ceci entraîne la nécessité d'interpoler les points de la signature avant tout processus de vérification. Evidemment, l'interpolation affecte la qualité du signal original de la signature car cette opération infère un décalage dans le temps entre le tracé fait à la main et sa réalisation sur l'écran tactile du PDA.

Par ailleurs, nous savons très bien que les conditions de capture peuvent altérer significativement la qualité du signal à l'acquisition et ainsi induire des distorsions dans le tracé de la signature. De ce fait, l'association de la tablette et du stylo est celle qui s'apparente le mieux à l'image traditionnelle de la feuille de papier et du crayon dans son utilisation. En effet, à l'inverse des PDAs par exemple, où l'utilisateur ne doit pas poser sa main sur l'écran tactile, l'utilisateur peut poser sa main sur la surface de la tablette sans affecter le processus d'acquisition. Ainsi, la tablette graphique facilite la procédure d'acquisition à l'utilisateur tout en lui assurant une écriture naturelle.

Par conséquent, dans la suite de notre travail, la tablette graphique sera considérée comme une "plateforme fixe" utilisée dans un but d'effectuer des acquisitions dans des conditions contrôlées. Alors que le PDA sera considéré comme une "plateforme mobile" utilisée pour acquérir des signatures dans des conditions dégradées (en mobilité).

1.4.2 Variabilité de la signature en-ligne

1.4.2.1 Variabilité temporelle

La signature est une modalité biométrique purement comportementale liée à un geste spécifique et unique pour une personne. Cependant, signer est un geste

rapide et répétitif, qui n'induit pas à la même signature à chaque fois. Cette variabilité est communément appelée "variabilité intra-classe". Par conséquent, les signatures successives d'une même personne varient globalement et localement, et sont donc similaires mais pas identiques. L'impact de cette variabilité intra-classe diffère d'une personne à une autre ; pour certaines personnes ayant une signature trop instable, il est difficile d'établir un modèle représentatif de leur signature.

Par ailleurs, la signature présente une variabilité au cours du temps : elle évolue au cours de la vie, au fil des évènements, des actions et des moments. Elle peut être fortement altérée par l'âge ou la maladie. De plus, une grande variabilité peut être observée dans les signatures en fonction de l'état psychologique et émotionnel de la personne.

On considère souvent pour une personne donnée plusieurs réalisations de sa signature pour pouvoir prendre en compte les variations de sa signature. Cependant, si cette personne présente une forte variabilité intra-classe, elle peut mettre en échec le système de vérification. En fait, plus la variabilité tolérée sera grande, plus le système va rejeter les signatures authentiques, et plus la probabilité qu'une imitation soit acceptée sera grande.

1.4.2.2 Variabilité du style

En graphologie, la signature manuscrite est considérée comme étant l'expression de notre propre personnalité, dans le sens où l'on ne choisit pas sa façon de signer par hasard. On rencontre même certaines personnes qui signent différemment en fonction du contexte auquel elles sont confrontées.

La signature manuscrite est considérée aussi comme un "label social" qui dévoile notre rapport à la société étant donné que l'aspect de la signature varie en fonction des pays et des habitudes culturelles. Par exemple, les signatures asiatiques sont constituées de traits courts, séparés par des levés de stylo et orientés suivant l'axe horizontal ou vertical, tandis qu'en arabe, le sens d'écriture va de droite à gauche, la plupart des lettres s'attachent entre elles, et certaines sont associées à des points souscrits ou suscrits.

De plus, chaque personne essaye de personnaliser sa signature pour la rendre unique. Ainsi, la manière de signer varie d'une personne à une autre révélant différentes tendances graphologiques : la signature peut être simple ou compliquée,

présentant une inclinaison ascendante ou descendante ; elle peut être composée de lettres lisibles ou de fioritures.

Ces différences sont nettement observées dans les signatures occidentales : certaines personnes ont des signatures informes ressemblant à des paraphes qui se présentent sous forme de graffiti grossiers (voir Figure 1.4.a), comme dans le cas des signatures européennes ; d'autres signent en apposant leurs propres nom et prénom, et cela de manière cursive (voir Figure 1.4.b), parfois calligraphiée, associée souvent à des effets de styles (courbes, traits, etc.), comme dans le cas des signatures anglo-saxonnes.

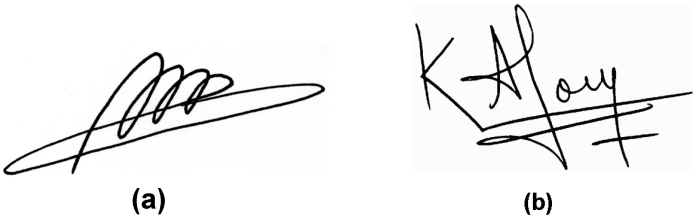

FIGURE 1.4: Exemple d'une signature (a) paraphe et d'une signature (b) cursive.

Toutes ces différences de styles compliquent la vérification de l'identité par la signature manuscrite, à l'inverse des autres modalités physiologiques où les données biométriques suivent un format bien spécifique. Effectivement, un système de vérification par signature manuscrite ne se comportera pas de la même manière sur les signatures "paraphes" comme sur les signatures "cursives". Il serait donc intéressant de classer au préalable les signatures de la base en groupes suivant le style d'écriture (paraphes, cursives,...), et leur appliquer ensuite un traitement spécifique en fonction du type de signature.

1.4.3 Vulnérabilité de la signature en-ligne aux attaques

Bien qu'il semble difficile de voler une caractéristique biométrique d'un individu, il est cependant facile de contourner un système de vérification en utilisant des données biométriques falsifiées.

La signature manuscrite est une modalité particulière qui diffère des autres modalités biométriques car son utilisation répandue depuis très longtemps a fortement motivé les attaques. En fait, il est possible de produire une fausse empreinte digitale ou d'imiter la voix de quelqu'un mais cela fait appel à des procédés très sophistiqués. Par contre, n'importe qui peut imiter une signature en disposant

d'une image de cette dernière. Si l'imitateur s'entraîne suffisamment bien, il arrive à reproduire des imitations assez similaires aux signatures originales, ce qui pourra tromper le système de vérification, lorsque la signature est statique et non pas dynamique.

Pour tester la capacité d'un système de vérification de signature en-ligne à discriminer les signatures authentiques des imitations, on utilise souvent des bases de signatures avec différents types d'imitations. Il existe essentiellement trois types d'imitations de la signature en-ligne. Les imitations dites "statiques", imitant seulement l'image de la signature. Les imitateurs possèdent l'image de la signature originale sur papier et essaient de la reproduire. Les imitations dites "dynamiques", imitant l'image de la signature ainsi que le geste du signataire. Dans ce cas, des enregistrements de vidéos de la personne en train de signer ou un logiciel d'affichage sont souvent fournis aux imitateurs pour revoir la signature en question en train d'être réalisée afin de pratiquer le geste, avant d'imiter. Les imitations dites "aléatoires" n'ont aucune information sur la signature à imiter. Ce dernier type d'imitations est aussi appelé imitations à "zéro effort". Comme son nom l'indique, on utilise souvent dans ce dernier cas les signatures des autres personnes comme imitations.

Il est toutefois évident que la réalisation d'une bonne imitation "dynamique" requiert beaucoup d'entraînements car il est très difficile, voire impossible, de reproduire à la fois la forme et la dynamique de la signature.

Par ailleurs, une bonne imitation dépend non seulement de la capacité d'un imposteur à imiter une signature et des informations dont il avait connaissance sur la signature originale, mais aussi de la difficulté de la reproduction de cette dernière. Manifestement, une signature stable avec beaucoup de changements de direction et de rythme sera beaucoup plus difficile à reproduire qu'une signature relativement variable constituée d'un tracé simple et sans changement de rythme.

Jusqu'à présent, nous avons exposé les problèmes associés à l'utilisation de la signature en-ligne comme moyen de vérification de l'identité. Ces problèmes affectent fortement les performances des systèmes de vérification, et motivent donc la nécessité d'étudier la qualité des signatures. En effet, la qualité des signatures dépend des conditions d'acquisition et de l'utilisateur lui-même (variabilité intra-classe et style d'écriture). De plus, les imitations peuvent être, selon leur similarité aux signatures originales, de bonne ou de mauvaise qualité, et influencent ainsi différemment la fiabilité des systèmes de vérification.

Dans la section suivante, nous nous attacherons à mettre en évidence la démarche que nous avons adopté dans le cadre de l'étude des mesures de qualité des signatures en-ligne, ainsi que les contributions majeures de notre travail de thèse.

1.5 Contributions majeures de cette thèse

Cette thèse a pour principal objectif l'introduction de nouvelles mesures permettant de quantifier la qualité des signatures en-ligne, en vue d'établir des critères automatiques de fiabilité des systèmes de vérification.

Comme les systèmes de vérification sont évalués sur des bases de données contenant des signatures authentiques et des imitations, leur fiabilité, en l'occurrence leur capacité à séparer les signatures authentiques des imitations, dépend non seulement de la qualité des signatures authentiques, mais aussi de celle des imitations.

Ainsi, afin d'atteindre notre objectif, nous avons adopté une démarche articulée sur trois niveaux :

1 - quantifier la qualité des signatures authentiques,

2 - quantifier la vulnérabilité d'une signature authentique exposée aux imitations,

3 - quantifier la qualité des imitations.

Quantifier la qualité des signatures authentiques Dans ce premier niveau, abordé dans le Chapitre 4, seules les signatures authentiques sont prises en compte. Nous proposons de quantifier la qualité de ces signatures en utilisant une nouvelle mesure basée sur le concept d'entropie [18], dénommée "*Entropie Personnelle*". Cette nouvelle mesure de qualité permet de classer automatiquement les personnes en 3 catégories cohérentes en terme d'aspect visuel, de complexité et de variabilité. Elle permet aussi de détecter les personnes problématiques, celles qui sont très difficiles à reconnaître et à caractériser, ainsi que les personnes les plus fiables en terme de performances.

L'originalité de notre nouvelle mesure de qualité réside dans les aspects suivants :

a - Nous avons proposé une définition objective et concrète de ce qu'est la qualité d'une signature, en l'identifiant comme étant le contenu d'information véhiculée par cette signature.

b - Nous avons alors proposé une mesure de qualité quantitative, indépendante de tout classifieur, basée sur le concept d'entropie, très utilisé dans la théorie de l'information.

c - Nous avons calculé l'*Entropie Personnelle* de manière très originale et adaptée au tracé de la signature : nous avons estimé les densités de probabilité localement, sur des portions de la signature, par le biais des Modèles de Markov Cachés. Ce dernier est entraîné sur un ensemble de signatures authentiques d'une personne donnée.

d - Notre nouvelle mesure de qualité englobe, en une seule mesure, les trois critères habituels de la littérature, à savoir, la complexité, la variabilité et la lisibilité.

e - Notre nouvelle mesure de qualité s'est avérée très utile pour un grand nombre d'applications liées à la vérification d'identité par la signature en-ligne, décrites dans le Chapitre 7.

Ce travail a conduit à la publication de 4 articles de conférences [42, 49, 50, 25] et un article de revue [43] (voir Annexe C).

Quantifier la vulnérabilité d'une signature authentique exposée aux imitations Dans ce deuxième niveau, abordé dans le Chapitre 5, les signatures authentiques ainsi que les imitations sont prises en compte. Quantifier la vulnérabilité d'une signature aux attaques est un concept original, très peu étudié dans la littérature. Il permet d'associer chaque personne à un certain niveau de fiabilité de sa signature.

Nous proposons dans le Chapitre 5 de quantifier la vulnérabilité en utilisant le concept d'entropie relative, appelée aussi la divergence de Kullback-Leibler [18]. La mesure de vulnérabilité proposée est appelée "*Entropie Relative Personnelle*". La vulnérabilité associée à chaque personne est calculée comme étant la distance de Kullback-Leibler entre les distributions de probabilité locales estimées sur les signatures authentiques de la personne et celles estimées sur les imitations qui lui sont associées. Et ce, par le biais de deux Modèles de Markov Cachés, où l'un est appris sur les signatures authentiques de la personne et l'autre sur les imitations

associées à cette personne. Plus la distance de Kullback-Leibler est faible, plus la personne est considérée comme vulnérable aux attaques.

Confrontée à plusieurs types d'imitations, notre mesure d'Entropie Relative Personnelle permet de sélectionner plus finement les personnes les plus fiables en terme de performance. Cette mesure est plus appropriée à notre problématique qui est l'évaluation de la fiabilité des systèmes de vérification, car elle intègre simultanément non seulement les trois critères usuels de la littérature (complexité, variabilité et lisibilité) comme notre mesure d'Entropie Personnelle calculée précédemment, mais elle intègre aussi le critère de vulnérabilité aux attaques.

Ce travail a conduit à la rédaction d'un article soumis à la revue *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Annexe C).

Quantifier la qualité des imitations Dans ce troisième niveau, développé dans le Chapitre 6, nous nous intéressons à un nouveau concept original qui n'a jamais été abordé auparavant dans la littérature, à savoir mesurer la qualité des imitations. En effet, jusqu'à présent les imitations se voyaient octroyées la mention "bonne" ou "mauvaise" imitation de manière subjective ! Alors que dans ce travail, nous proposons d'élargir la notion de mesure de qualité aux imitations, et de leur attribuer ainsi de manière quantitative l'étiquette "bonne" ou "mauvaise" imitation. Notons que mesurer la qualité des imitations d'une base de signatures a pour but d'apporter un niveau de confiance vis-à-vis de la validation des résultats obtenus après évaluation des systèmes de vérification sur cette base.

Dans ce troisième niveau, les signatures authentiques ainsi que les imitations sont prises en compte. Notre nouvelle mesure de qualité des imitations est une extension de la mesure d'*Entropie Personnelle* calculée dans le premier niveau, et qui est adaptée au contexte des imitations : elle est définie comme étant la dissimilarité existant entre l'entropie associée à la personne à imiter et celle associée à l'imitation. Confrontée à différentes bases de signatures contenant plusieurs types d'imitations (statiques, dynamiques et professionnelles), notre mesure de qualité permet de quantifier la qualité de chacun des types d'imitations, et cela indépendamment de tout classifieur.

Ce travail a conduit à la rédaction d'un article soumis à la revue *Pattern Recognition* (Annexe C).

Les différentes applications de la mesure d'Entropie Personnelle Dans cette partie, développée dans le Chapitre 7, nous confirmons d'abord la validité de la mesure d'*Entropie Personnelle* en la confrontant aux performances de plusieurs systèmes de vérification soumis lors de la campagne d'Evaluation BioSecure de signatures en-ligne "BSEC'2009" [25, 52], organisée par notre équipe à Télécom SudParis. Puis, nous nous intéressons à l'intérêt applicatif potentiel de notre mesure d'*Entropie Personnelle*. Nous montrons qu'elle ne sert pas uniquement à classer les personnes suivant la qualité de leur signature, mais contribue aussi à l'amélioration des performances des systèmes de vérification de différentes manières. Nous verrons qu'elle peut être utilisée pour :

- améliorer la procédure d'enregistrement [43].

- quantifier la dégradation de la qualité des signatures due au changement de plateforme.

- sélectionner les meilleures signatures de référence pour chaque personne.

- détecter les signatures aberrantes (des anomalies) pouvant introduire des erreurs importantes lors de la vérification.

- quantifier la pertinence de certains paramètres dans le contexte de variabilité temporelle [50].

1.6 Organisation de la thèse

Dans ce chapitre, nous avons défini le contexte général et les motivations de cette thèse. Nous avons aussi donné brièvement la ligne directrice de notre travail et cité les contributions majeures de cette thèse.

Dans le Chapitre 2, nous présenterons les notions de base d'un système de vérification de l'identité par la signature en-ligne. Nous décrirons aussi les principales approches de classification les plus utilisées dans la littérature qui vont servir à nos expérimentations.

Dans le Chapitre 3, nous dresserons un état des lieux concernant les différents travaux de la littérature relatifs aux mesures de qualité des signatures manuscrites.

Le Chapitre 4 sera dédié à la quantification de la qualité des signatures authentiques par la mesure d'Entropie Personnelle.

Le Chapitre 5 présentera la mesure d'Entropie Relative Personnelle utilisée pour la quantification de la vulnérabilité des signatures authentiques aux attaques.

Le Chapitre 6 sera consacré à la quantification de la qualité des imitations en exploitant la mesure d'Entropie Personnelle.

Dans le Chapitre 7, nous développerons les différentes applications de l'Entropie Personnelle.

Enfin, le Chapitre 8 conclura les travaux de cette thèse et ouvrira les nouvelles perspectives.

Chapitre 2

Vérification d'identité par la signature en-ligne

En général, les systèmes biométriques peuvent être utilisés selon deux modes différents : la vérification ou l'identification de personnes. On parle d'identification lorsque l'identité de l'utilisateur est inconnue. Elle permet de savoir si l'individu est présent dans une base de données. Dans ce cas, la donnée biométrique de l'individu en question est comparée à toutes celles stockées dans la base de données afin de déterminer l'identité de la personne.

On parle de vérification d'identité lorsque la personne clame son identité avant tout usage de la biométrie. Dans ce cas, le système compare la caractéristique biométrique acquise sur le capteur au moment du test à celle servant de référence pour l'identité proclamée ; ainsi, le système infirme on confirme l'identité proclamée.

Cette thèse rentre dans le cadre de la vérification de l'identité par la signature en-ligne. Toutefois, il est important de rappeler que notre travail ne porte en aucun cas sur l'élaboration et la mise en œuvre de nouveaux systèmes de vérification de signature. En revanche, nous utiliserons des systèmes de vérification de signature en-ligne existants pour évaluer les nouvelles mesures de qualité proposées dans ce travail sur des bases de signatures à notre disposition.

Ce chapitre est organisé en deux grandes parties. La première partie comportera des généralités sur le fonctionnement des systèmes de vérification et l'évaluation de leur performance. La deuxième partie sera consacrée à présenter les bases de

signatures dont nous disposons, ainsi que les approches de classifications sur lesquelles reposent les systèmes de vérifications utilisés au cours de notre travail.

2.1 Principe d'un système de vérification

Comme pour toute modalité biométrique, la structure d'un système de vérification d'identité par signature en-ligne comprend deux phases distinctes : l'enregistrement et le test (la vérification). En général, un système de vérification comporte cinq modules dont certains sont communs aux deux phases : module d'acquisition, module d'extraction de caractéristiques, module de modélisation, module de comparaison et module de décision. La Figure 2.1 illustre la structure générale et les étapes d'un système de vérification d'identité par la signature en-ligne.

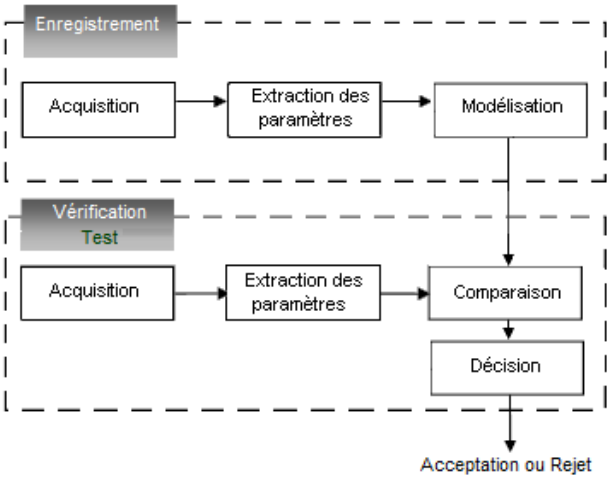

FIGURE 2.1: Etapes d'un système de vérification d'identité.

Les modules d'acquisition et d'extraction de caractéristiques sont communs à la phase d'enregistrement et de test. Le module d'acquisition consiste à acquérir les signatures de référence de l'utilisateur par le biais d'un capteur digital. Actuellement et depuis la première compétition internationale de systèmes de vérification de signature en-ligne en 2004 (SVC'2004) [122], le nombre de signatures d'enregistrement usuel est de cinq. Comme nous l'avons déjà évoqué dans le Chapitre 1, cette étape est très importante car elle affecte tout les processus de reconnaissance ultérieurs. En effet, la qualité de ces données d'enregistrement a été montrée influant particulièrement sur les performances des traitements ultérieurs de la phase de vérification (voir Figure 2.1). Lors de cette étape d'acquisition, la signature est représentée sous forme d'une séquence de fonctions temporelles brutes. Dans le cas

d'une tablette graphique par exemple, cinq fonctions temporelles sont acquises : les coordonnées *(x,y)*, la pression ainsi que les angles d'inclinaison du stylo (Azimuth et Altitude).

Le module d'extraction prend en entrée les signatures acquises par le module de capture et extrait à partir des fonctions temporelles d'autres caractéristiques afin de former une nouvelle représentation de la signature. Idéalement, cette nouvelle représentation est censée être unique pour chaque personne et relativement invariante aux variations intra-classe. Au terme de cette étape, la signature sera représentée soit sous forme d'une séquence de vecteurs contenant différentes caractéristiques temporelles (vitesse, accélération, etc.), soit sous forme d'un seul vecteur contenant des caractéristiques globales comme la durée totale et la vitesse moyenne calculée sur toute la signature.

A l'enregistrement, le vecteur des caractéristiques extrait de la signature est appelé référence et est stocké dans une base de données. Tandis qu'à la phase de vérification, les modules d'acquisition et d'extraction de caractéristiques permettent d'obtenir une représentation de la signature à tester dans l'espace des caractéristiques.

Après l'acquisition et l'extraction des caractéristiques, selon les approches, soit on construit par apprentissage un modèle de la signature d'une personne et ce modèle est appris sur les signatures d'enregistrement (base d'apprentissage du modèle dans ce cas) ; soit on utilise toutes les signatures d'enregistrement en tant que signatures de référence de la personne (base de référence dans ce cas).

Les module de comparaison et de décision sont utilisés seulement dans la phase de test. A la phase de vérification (ou de test), l'utilisateur proclame son identité. Les paramètres sont extraits de la signature de test qui est présentée au système. Selon les approches, la signature de test est soit présentée en entrée du modèle correspondant à l'identité proclamée, soit comparée avec les signatures de référence associées à l'identité proclamée. Après cette étape, le module de décision rejette ou accepte la signature de test en fonction du score obtenu à la sortie du module de comparaison. Ce score correspond à une valeur de similarité (ou de dissimilarité) entre les caractéristiques extraites de la signature test et celles des signatures stockées (signatures de référence).

2.2 Mesure des performances d'un système de vérification

Afin de décider si une signature de test est authentique ou pas, le système compare le score de l'échantillon de test à un seuil de décision. Si ce score est un score de similarité, s'il dépasse le seuil, alors le système accepte l'identité proclamée, sinon il la rejette (voir Figure 2.2). Inversement, dans le cas où le score est un score de dissimilarité, si ce dernier est inférieur au seuil alors le système accepte l'identité proclamée, sinon il la rejette.

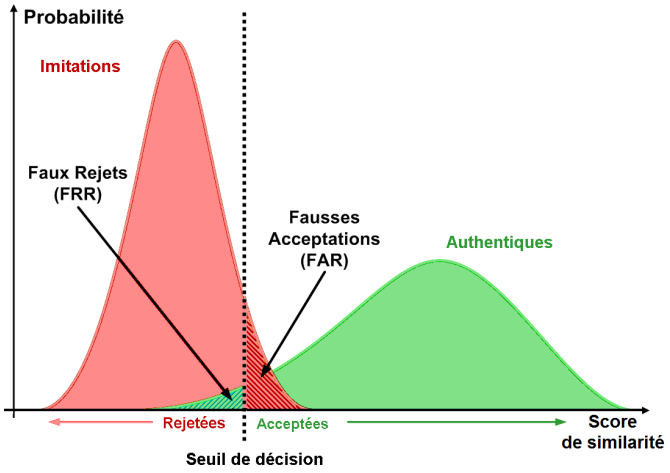

FIGURE 2.2: Distribution des scores des signatures authentiques et des imitations.

Lorsqu'un système fonctionne en mode vérification, celui-ci peut faire deux types d'erreurs. Il peut rejeter un utilisateur légitime et dans ce premier cas on parle de faux rejet. Il peut aussi accepter malencontreusement un imposteur et on parle dans ce second cas de fausse acceptation. La performance d'un système se mesure donc en se basant sur son taux de faux rejet (*False Rejection Rate* ou FRR) et son taux de fausse acceptation (*False Acceptance Rate* ou FAR) [95]. La Figure 2.2 illustre le FRR et le FAR à partir des distributions des scores authentiques et imposteurs.

Ainsi, pour une valeur du seuil quelconque, nous calculons les taux d'erreur correspondant à ces deux types d'erreur. Le taux de faux rejets est le pourcentage des données de test authentiques qui ont été rejetées :

$$FRR = 100 * \frac{nombre\ de\ tests\ authentiques\ rejetés}{nombre\ total\ de\ tests\ authentiques}(en\%)$$

De même, le taux de fausses acceptations est le pourcentage des imitations qui ont été acceptées :

$$FAR = 100 * \frac{nombre\ d'imitations\ acceptées}{nombre\ total\ d'imitations}(en\%)$$

Le taux de faux rejets et celui de fausses acceptations dépend du seuil de sécurité, et sont inversement proportionnels. Plus la valeur du seuil sera grande, plus il y aura de faux rejets et moins de fausses acceptations, et inversement, plus la valeur du seuil sera petite, moins il y a aura de faux rejets et plus de fausses acceptations. Le choix de la valeur du seuil à utiliser dépend principalement de la finalité du système de vérification. Cette valeur est choisie de manière à faire un compromis adéquat entre la sécurité et l'utilité du système.

Souvent, on caractérise un système de vérification de signature en-ligne par un point où les deux courbes, *FAR* en fonction du seuil et *FRR* en fonction du seuil, se croisent. En ce point, les taux d'erreur *FAR* et *FRR* sont égaux, ils sont donc représentés par une valeur unique qui est le taux d'erreur égal (*Equal Error Rate* ou *EER*). En fait : $EER = FAR = FRR$.

La valeur du *EER* est un indicateur de performance d'un système de vérification, indépendamment des exigences d'une application spécifique (rapport entre *FRR* et *FAR*). Plus cette valeur est faible, meilleur est le système. L'inconvénient de l'*EER* est qu'il ne nous permet de comparer les systèmes qu'en un seul point de fonctionnement, et donc ne permet pas d'évaluer le système à des niveaux de sécurité différents (faible *FAR* par exemple).

Afin d'évaluer un système de vérification indépendamment du seuil, nous faisons varier la valeur du seuil intrinsèquement sur un intervalle donné, et pour chaque valeur du seuil, nous calculons le taux de fausses acceptations et le taux de faux rejets.

Pour obtenir une représentation compacte des performances d'un système de vérification au travers d'une seule courbe, nous utilisons souvent les courbes *DET* (*Detection Error Tradeoff* [79]) pour représenter les *FRR* en fonction des *FAR*, comme illustré sur la Figure 2.3. L'échelle est basée sur une distribution normale pour rendre la courbe plus lisible et plus exploitable. Cette représentation graphique est très utilisée pour comparer différents systèmes de vérification qui ont des performances similaires.

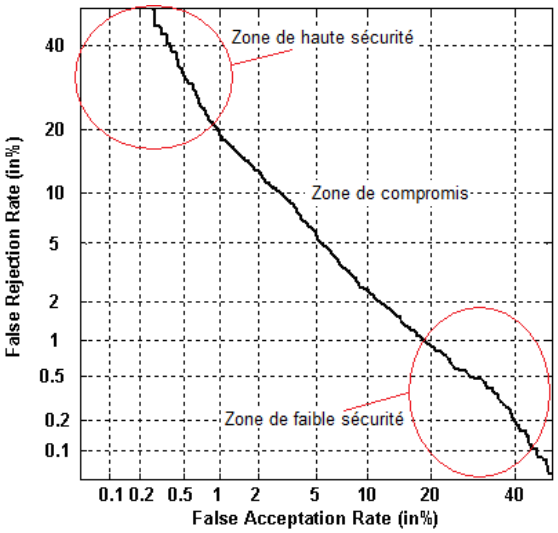

FIGURE 2.3: Exemple de courbe DET.

2.3 Calcul de l'intervalle de confiance

Les bases de signatures utilisées pour évaluer les performances des systèmes de vérification automatiques sont souvent de tailles réduites, et ne sont jamais totalement représentatives de la population globale. Ainsi, l'usage du taux d'erreur *EER* seul pour évaluer les performances des systèmes de vérification ne permet pas d'apporter un niveau de confiance en ce qui concerne la généralisation des résultats sur une plus grande base. Par ailleurs, les protocoles d'évaluation exigent la division de la base en un ensemble de référence et un ensemble de test. Ainsi, les performances obtenues sont dépendantes du découpage référence-test. Par conséquent, il est nécessaire de calculer un intervalle de confiance à l'*EER* afin d'indiquer la confiance que l'on peut avoir dans les résultats obtenus.

Une manière de calculer un intervalle de confiance est d'utiliser la méthode de rééchantillonnage *bootstrap* [73, 105]. Il s'agit d'une technique qui permet de faire de l'inférence statistique sur de nouveaux échantillons tirés à partir de l'échantillon initial. Cette méthode permet de construire des intervalles de confiance quand on a des hypothèses faibles, ou aucune hypothèse n'est fait au sujet de la distribution fondamentale des données.

Disposant d'un échantillon initial de taille N destiné à donner une certaine information sur une population, on tire au hasard avec remise, parmi cet échantillon initial, de nouveaux sous-ensembles de même taille n. Et on répète cette opération

k fois, où k est grand. On analyse ensuite les nouvelles observations ainsi obtenues après chaque rééchantillonnage pour affiner l'inférence faite sur l'observation initiale.

Suivant ce principe, pour estimer l'intervalle de confiance à l'*EER*, on effectue des tirages aléatoires avec remise de n=5 signatures parmi les N signatures authentiques de la personne disponibles dans la base de données. Ainsi, à chaque tirage k, on obtient un nouvel ensemble de référence pour chaque personne, et chaque signature peut être retirée plus d'une fois, ou pas du tout. Par la suite, pour chaque tirage k, on réévalue de la même façon un nouveau taux d'erreur égal EER_k, en considérant dans l'ensemble de test les signatures authentiques restantes ainsi que les imitations. Nous pouvons de la sorte estimer la variation du taux d'erreur sur tous les ensembles rééchantillonnés.

D'après la loi des grands nombres, lorsque k tend vers l'infini, la variable qu'on veut estimer tend vers une variable normale. Donc, lorsque le nombre de tirage est relativement grand, la distribution tend à suivre une loi normale, et l'intervalle de confiance peut être déterminé grâce aux percentiles de la distribution normale [105].

L'intervalle de confiance à 95% est défini alors comme suit :

$$IC = EER \pm 1,96 * \frac{\sigma}{\sqrt{k}}$$

où *EER* est le taux d'erreur global estimé sur tous les tirages confondus, k est le nombre de tirages, et σ est la variance des k taux d'erreur calculés sur les différents tirages k.

L'intervalle de confiance représente une mesure de confiance des résultats obtenus. Plus il est faible, plus les résultats sont fiables. L'intervalle de confiance à 95% indique que la probabilité de se tromper lorsqu'on affirme que l'*EER* se situe à l'intérieur de cet intervalle est de 0,05. La valeur de 1,96 correspond au percentile de 0,05.

Evidemment, plus le nombre de tirages k est grand, plus la statistique est précise, surtout dans les cas où les données sont fortement variables. Néanmoins, plus k est grand, plus les expériences deviennent lourdes en terme de temps de calcul. Nos expériences ont montré qu'un nombre de tirages de 20 est généralement suffisant pour une bonne estimation des valeurs de la moyenne et l'écart-type des erreurs. En plus, au delà de 20 tirages, nous avons observé une stabilité de ces deux valeurs.

2.4 Description des bases de signatures utilisées

Nous disposons de cinq bases différentes de signatures en-ligne : Biomet [39], MCYT-100 [91], Philips [23, 24], et BioSecure DS2 et DS3 [54, 90]. Ces bases ont été acquises avec différents capteurs et suivant différents protocoles. Nous décrirons ces bases de signature en détail dans ce qui suit.

2.4.1 La base Biomet

Biomet est une base de données multimodale [39] comprenant entre autre la signature manuscrite en-ligne.

Les signatures dans cette base sont échantillonnées par la tablette Wacom Intuos 2 A6 qui a une surface active d'acquisition de 127 x 106 mm^2. La fréquence d'acquisition est de 100 Hz. Bien que la résolution originale de la tablette soit 12700 x 9650 pixels sur toute la surface active, le programme d'acquisition l'a tronquée pour acquérir les coordonnées à une résolution plus faible, de seulement 1024 x 768 pixels (environ 5840 pixels par cm^2). En chaque point de la signature, le programme d'acquisition capture aussi la pression et l'inclinaison du stylo, par le biais des angles d'Azimut et d'Altitude du stylo. La pression est quantifiée en 1024 valeurs différentes qui varient de 0 à 1023. Quant aux angles d'inclinaison, comme le montre la Figure 2.4, l'angle d'Azimut est codé avec 360 valeurs {0 , 10 , 20 , ... , 3590} qui correspondent respectivement à {0°, 1°, 2°, ... , 359°}. L'angle d'Altitude est codé de la même façon, mais les valeurs varient de 300 à 900 par intervalle de 10 car la tablette Wacom ne peut mesurer que l'altitude entre 30° et 90°.

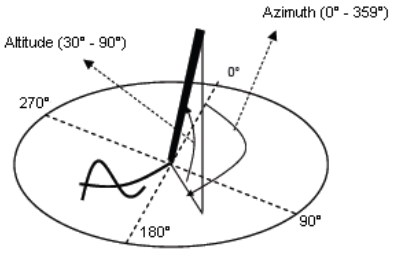

FIGURE 2.4: Angles d'Azimut et d'Altitude utilisés pour représenter la position du stylo en 3 dimensions dans la base Biomet.

Biomet contient des signatures de 84 personnes. Les signatures d'une personne sont acquises en deux sessions espacées de 5 mois. Le but étant d'étudier la variabilité dans le temps de la signature manuscrite, autrement dit la dégradation

des performances du système de vérification au cours du temps. La Figure 2.5 montre pour une personne donnée, un exemple de ses signatures authentiques acquises durant la première session (Figure 2.5.a) et la deuxième session (Figure 2.5.b), ainsi que ses imitations acquises en chaque session (Figure 2.5.c).

FIGURE 2.5: Exemples de signatures d'une personne appartenant à la base Biomet : (a) Signatures authentiques de la Session 1 (b) Signatures authentiques de la session 2 (c) imitations.

Lors de la première session de la base Biomet, 5 signatures authentiques et 6 signatures imitées ont été acquises par personne. Lors de la deuxième session, 10 signatures authentiques et 6 signatures imitées ont été saisies par personne. Les 12 signatures imitées résultantes d'une personne ont été réalisées par 4 imitateurs différents, chacun en a fait 3. Les imposteurs ont essayé d'imiter seulement l'image de la signature, et cela sans entraînement.

2.4.2 La base MCYT-100

La création de cette base a été motivée par le manque de grandes bases biométriques publiques utilisables pour l'évaluation des performances des systèmes de reconnaissance. Dans ce contexte, le projet espagnol MCYT [91], achevé à la fin de 2003, avait pour mission l'acquisition d'une base biométrique bimodale, contenant les empreintes digitales et les signatures en-ligne de plus de 300 personnes.

Dans cette base, la signature est échantillonnée à 100 Hz par la tablette à digitaliser Wacom Intuos A6 USB. Cinq paramètres sont acquis pour chaque échantillon de signature : les coordonnées *(x,y)*, la pression et les angles d'inclinaison du stylo (Azimut et Altitude). La résolution de la pression et des angles

d'inclinaison du stylo est la même que dans la base Biomet. Par contre, les coordonnées sont acquises à une plus haute résolution qui est la résolution originale de la tablette (10^6 pixels par cm^2).

La Figure 2.6 montre des exemples de signatures authentiques et imitées de trois personnes différentes : les deux signatures sur la gauche sont authentiques, et celle sur la droite est une imitation.

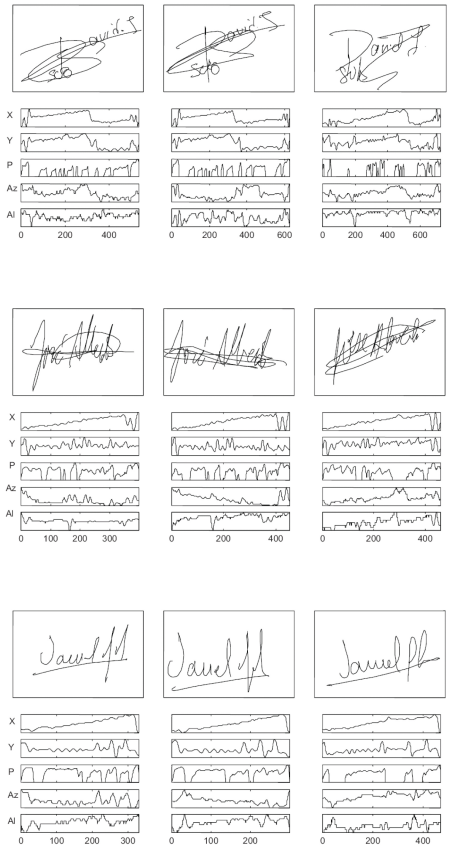

FIGURE 2.6: Signatures de trois personnes différentes appartenant à la base MCYT-100.

La base MCYT complète contient 330 signataires. Par contre, nous ne disposons que d'une sous-partie de la base qui est gratuite, dénommée la base MCYT-100, contenant les 100 premiers signataires. Chaque personne a 25 signatures authentiques et 25 imitations. Afin d'étudier la variabilité de la signature à court-terme, chaque personne a donné ses signatures authentiques en 5 petites sessions de 5 signatures. La procédure d'acquisition est comme suit : chaque personne n a fait 5 signatures authentiques, puis 5 imitations de la personne $n-1$. Cette procédure est répétée 4 fois en imitant les personnes $n-2$, $n-3$, $n-4$ et $n-5$. Les imposteurs ont imité seulement l'image de la signature en essayant de signer le plus

naturellement possible et sans artefacts (comme les pauses et les ralentissements par exemple).

2.4.3 La base Philips

Les signatures dans cette base [23, 24] sont acquises par la tablette à digitaliser PAID à la fréquence de 120 Hz. En chaque point échantillonné de la signature, la tablette enregistre 5 paramètres : les coordonnées *(x,y)*, la pression du stylo et l'inclinaison du stylo par rapport aux deux axes x et y (θx et θy) (voir Figure 2.7). La pression et les angles d'inclinaison sont enregistrés en 64 niveaux différents.

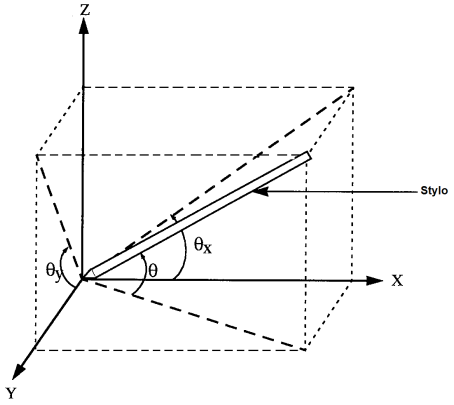

FIGURE 2.7: Les angles θx et θy utilisés pour définir la position spatiale du stylo dans la base Philips.

La base de signatures en-ligne Philips est très intéressante pour l'évaluation des systèmes de vérification à cause de sa richesse en types d'imitations. Elle permet d'étudier la capacité des systèmes à détecter des imitations issues de différents protocoles d'acquisition.

En effet, différents scénarios d'imitation ont été proposés : Le premier type d'imitations (SH - Over the-Shoulder- par dessus l'épaule) concerne les imposteurs qui imitent les informations dynamiques de la signature comme la vitesse, l'inclinaison du stylo, la pression...après avoir observé la personne en train de signer. Le deuxième type d'imitations (HI - Home-Improved - améliorées chez soi) concerne les imposteurs ayant uniquement l'image statique de la signature à imiter. Ils peuvent s'entraîner à imiter l'image de cette signature. Le troisième type d'imitations (PR - PRofessional - professionnels) concerne les individus ayant une grande expérience professionnelle en expertise des manuscrits. En utilisant leur expérience, ils fournissent des imitations de très haute qualité sans entraînement

préalable. Toutefois, pour ce type d'imitations, il n'y a que 4 imitateurs et seule la forme de la signature est imitée et non pas la dynamique.

Cette base [23, 24] contient les signatures de 51 personnes dont 1530 signatures authentiques (30 par personne), 1470 SHs (30 par personne pour 49 personnes), 1530 HIs (30 par personne) et 200 PRs (10 par personne pour seulement 20 personnes).

2.4.4 Les bases BioSecure DS2 et DS3

BioSecure DS2 (Data Set 2) et DS3 (Data Set 3) sont deux bases multimodales contenant la signature en-ligne [54, 90]. Elles ont été acquises dans le cadre du réseau d'excellence européen BioSecure (*Biometrics for Secure Authentication*).

Ce réseau d'excellence d'une durée de 3 ans (du 1er Juin 2004 au 31 Septembre 2007) regroupait 30 partenaires issus de 17 pays européens. Un des événements majeurs de BioSecure a été l'acquisition multi-sites (dans 12 laboratoires de recherche différents) d'une base de données biométriques multimodale de taille importante avec plusieurs configurations [54]. En effet, la base DS2 a été acquise sur une plateforme fixe suivant un scénario supervisé, alors que la base DS3 a été acquise sur une plateforme mobile dans des conditions dégradées. La signature manuscrite en-ligne est l'une des modalités acquises dans DS2 et dans DS3.

La particularité de ces deux bases est qu'elles ont en commun un nombre important de personnes. Afin d'analyser l'impact des capteurs et des conditions d'acquisition sur les performances des systèmes de vérification, on a demandé à ces personnes d'acquérir leurs signatures d'une part sur une tablette fixe (DS2), et d'autre part sur une tablette mobile (DS3). Dans la suite, nous décrivons ces deux bases séparément.

2.4.4.1 La base BioSecure DS2

BioSecure DS2 est une base multimodale qui contient la signature en-ligne. La partie de cette base concernant la signature en-ligne contient les données de 667 personnes acquises sur une plateforme fixe (tablette graphique). La base DS2 complète n'est pas publique. Cependant, une sous-partie de cette base contenant les signatures de 210 personnes est maintenant disponible, dénommée DS2-210 [54, 90].

La signature dans cette base est échantillonnée par la tablette graphique Wacom Intuos 3 A6 qui a une surface active d'acquisition de 270 x 216 mm^2. La résolution de la tablette est de 5080 lignes par pouce, la précision est de 0,25 mm et la hauteur maximale de détection est de 13 mm. La fréquence d'acquisition est de 100 Hz. En chaque point de la signature, 5 fonctions temporelles sont acquises : les coordonnées *(x,y)* du stylo, la pression du stylo (quantifiée en 1024 valeurs différentes) et les angles d'inclinaison (Azimut et Altitude).

Cette base a été acquise sur 2 sessions séparées de deux semaines. Chaque session contient 15 signatures authentiques et 20 imitations, acquises comme suit : chaque personne a donné ses signatures authentiques en 3 petites sessions de 5 signatures. Entre ces 3 sessions, la personne n a imité 5 fois la signature de deux autres personnes (les personnes $n-1$ et $n-2$ pour la session 1, et les personnes $n-3$ et $n-4$ pour la session 2), et cela après plusieurs minutes de pratique tout en ayant connaissance de la dynamique de la signature à imiter sur un écran d'ordinateur.

2.4.4.2 La base BioSecure DS3

BioSecure DS3 est aussi une base multimodale qui contient des signatures en-ligne. La partie de cette base concernant la signature en-ligne contient les données de 713 personnes acquises sur une plateforme mobile (PDA). Seulement une sous-partie de cette base contenant les signatures de 240 personnes est disponible [54, 90], dénommée DS3-240. Les 120 premières personnes de cette base sont communes à DS2-210.

Les signatures de cette base sont acquises sur le iPAQ hx2790 PDA HP, à la fréquence de 100 Hz et avec une résolution d'écran tactile de 1280 * 960 pixels. Trois fonctions temporelles sont capturées par le PDA : les coordonnées *(x,y)* et le temps écoulé entre deux points successifs. L'utilisateur signe en position debout en gardant le PDA dans sa main.

Cette base a été acquise sur deux sessions séparées de 6 semaines. Chaque session contient 15 signatures authentiques et 20 imitations, acquises suivant le même protocole que DS2 : chaque personne a donné ses signatures authentiques en 3 petites sessions de 5 signatures. Entre ces 3 sessions, la personne a imité 5 fois la signature de deux autres personnes (les personnes $n-1$ et $n-2$ pour la session 1, et les personnes $n-3$ et $n-4$ pour la session 2).

Afin d'obtenir des imitations de bonne qualité, les faussaires ont été amenés à imiter non seulement l'image de la signature mais aussi sa dynamique grâce à un logiciel de visualisation : le faussaire visualise sur l'écran du PDA la séquence d'écriture de la signature qu'il doit imiter, et peut même signer sur l'image de la signature afin d'obtenir une imitation de meilleure qualité, tant du point de vue de la dynamique que de la forme de la signature.

Dans notre travail de thèse, nous avons évalué les systèmes de vérification sur plusieurs sous-ensembles de la base BioSecure : les deux sous-ensembles DS2-104 et DS3-104 contenant les mêmes 104 personnes dont les signatures ont été acquises à Télécom sudParis, les deux sous-ensembles DS2-210 et DS3-240 disponibles au public et contenant les mêmes 120 premières personnes, et le sous-ensemble tenu séquestré DS2-382 contenant les signatures de 382 personnes.

2.5 Approches de classification adoptées

Il existe plusieurs approches pour la vérification de signatures manuscrites en-ligne. Dans cette partie, nous n'allons pas faire un état de l'art exhaustif de toutes les approches de vérification de signature en-ligne existantes, mais seulement signaler les grandes catégories de classifieurs que nous avons considérées dans notre travail. Plusieurs critères sont possibles pour classer les systèmes de vérification de signature en-ligne existants.

Un premier critère concerne la phase d'extraction de paramètres [82, 109]. En effet, comme évoqué brièvement dans la Section 2.1, on distingue dans ce contexte une approche dite "locale", où les caractéristiques de la signature sont extraites de façon plus ou moins locale : en chaque point [13, 34, 35, 46, 62, 67, 74, 72, 110, 87], ou en quelques points spécifiques [47, 114], ou en chaque portion du tracé de la signature [23, 24]. Dans ce cas, la signature est représentée par une séquence de vecteurs décrivant les propriétés locales de la signature. Comme les signatures d'une même personne peuvent varier en longueur, la taille des séquences de vecteurs diffère alors d'une signature à une autre.

On distingue aussi, une approche dite "globale", où les caractéristiques sont extraites sur la totalité de la signature [34, 48, 121, 69], décrivant ainsi la signature de manière plus globale, plus grossière. Dans ce cas, toutes les signatures sont représentées par des vecteurs de taille fixe, contenant des informations globales sur les signatures. Ces informations sont calculées par intégration des informations

locales donnant lieu à, par exemple, la vitesse moyenne, la durée totale, le rapport hauteur largeur de la signature, etc. Dans le cas de cette approche, on perd la notion de séquence temporelle, et la mise en correspondance entre deux signatures repose uniquement sur une distance, la distance euclidienne par exemple.

Dans la suite de notre travail, nous n'utiliserons pas l'approche globale pour étudier les performances des systèmes de vérification, et nous nous focaliserons uniquement sur les classifieurs basés sur l'approche locale. Notre choix est motivé par le fait que les performances des approches globales sont médiocres puisque les paramètres globaux contiennent peu d'information et ne sont pas assez discriminant [82, 109]. De plus, à l'inverse de l'approche locale, pour atteindre des performances acceptables, l'approche globale nécessite l'utilisation d'un nombre important de paramètres, de l'ordre de 50 à 100 [109]. De plus, l'approche globale n'est pas du tout adaptée à notre mesure de qualité proposée dans ce travail car cette dernière opère "localement" sur des portions de la signature.

Un deuxième critère possible pour classer les systèmes de vérification est celui de l'approche de classification adoptée [83, 44, 109]. Lorsqu'on cherche à comparer directement des signatures en-ligne constituées de séquences de vecteurs ordonnés dans le temps, la difficulté réside dans le fait que ces deux séquences ne sont pas forcément de même taille : l'une peut être plus longue que l'autre si elle a été tracée plus lentement. Ainsi, on distingue les principales méthodes de comparaison suivantes :

a) Celles basées sur des méthodes nécessitant un apprentissage comme les Modèles de Markov Cachés [82, 44, 96]. Dans le cadre de l'utilisation de cette approche, un individu est associé à un Modèle de Markov Caché donné, appris sur un ensemble de signatures d'apprentissage de cette personne (signatures servent de référence).

b) Celles basées sur un calcul de distance entre des séquences de tailles variables, comme la distance élastique [82, 44, 96].

Nous présentons dans ce qui suit les deux approches les plus utilisées dans l'état de l'art pour la vérification de signature en-ligne, à savoir : les Modèles de Markov Cachés (MMCs) et la distance élastique (Dynamic Time Warping ou DTW).

2.5.1 Modèles de Markov Cachés (MMCs)

La modélisation par les Modèles de Markov Cachés (MMCs) [96] est devenue la solution par excellence aux problèmes de reconnaissance des formes, mettant en œuvre des séquences de taille variable. Effectivement, ces modèles sont mathématiquement établis et s'appliquent à tout problème où l'information est incomplète, grâce au principe du maximum de vraisemblance. Leur formalisme probabiliste et leur capacité de modélisation des séquences en ont fait un instrument très efficace dans beaucoup d'applications. Ces méthodes s'imposent depuis plus de dix ans dans le domaine de la reconnaissance de la parole, puis de l'écriture, notamment dans la reconnaissance de la signature manuscrite en-ligne.

2.5.1.1 Structure générale d'un MMC

Un modèle de Markov Caché est un processus probabiliste doublement stochastique [96], qui se base sur l'existence d'un ensemble d'états cachés qui ne sont pas directement observables. Chaque état S_i définit une source possible pour une observation donnée O_j, et émet suivant sa propre distribution de probabilités des observations qui constituent la séquence effectivement observée $O = O_1, ..., O_T$ (voir Figure 2.8). La probabilité de chaque état ne dépend que de l'état qui le précède (hypothèse Markovienne). Les observations sont une fonction aléatoire de l'état qui, à son tour, change à chaque instant en fonction des probabilités de transition issues de l'état antérieur. Autrement dit, le passage d'un état à un autre s'effectue en tenant compte d'une probabilité de transition d'un état à un autre. La séquence d'observation est donc le résultat de deux processus stochastiques : la transition entre les états et les émissions des observations (voir Figure 2.8).

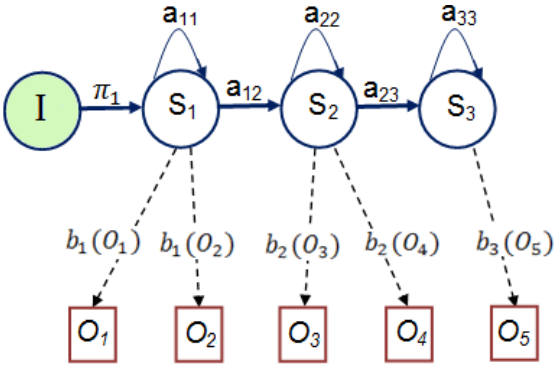

FIGURE 2.8: Paramètres d'un Modèle de Markov Caché. S : états, O : observations possibles, a : probabilités de transition, b : probabilités d'émission.

Ainsi, comme illustré sur la Figure 2.8, un modèle de Markov Caché est constitué de deux processus stochastiques, à savoir :

a) Un processus caché : une chaîne de Markov qui représente les états du modèle $S = S_1, S_2, S_3, ..., S_N$, cette chaîne vérifie la propriété de Markov à l'ordre 1 :

$$P(q_t = S_t | q_{t-1} = S_{t-1}, ..., q_{t-n} = S_{t-n}) = P(q_t = S_t | q_{t-1} = S_{t-1})$$

où q_t est l'état visité dans le modèle à l'instant t.

b) Un processus observable : une séquence d'observations $O = (O_1, O_2, ..., O_T)$.

Un Modèle de Markov Caché est caractérisé par les paramètres suivants :

- N : le nombre d'états du modèle : $S = S_1, S_2, S_3, ..., S_N$.

- A : la matrice de probabilites de transitions entre les états $A = \{a_{ij}\}$

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N$$

- π : la distribution initiale des états : $\pi = \{\pi_i\}$

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N$$

- B : La loi de probabilité d'émission de l'observation O_t dans l'état S_j :

$$b_j(O_t) = P(O_t | q_t = S_j), \quad 1 \leq j \leq N$$

o Si on considère un MMC discret, les observations appartiennent à un ensemble fini de symboles possibles, et la loi d'émission d'observation est une matrice de probabilité.

o Si on considère un MMC continu, ce qui est notre cas dans cette thèse, les observations appartiennent à un espace continu, et la loi d'émission d'observation est une loi paramétrique quelconque. Souvent, le mélange de gaussiennes [97] est utilisé pour modéliser la fonction de densité de probabilité dans chaque état :

$$b_j(O_t) = P(O_t | q_t = S_j) = \sum_{m=1}^{M} C_{jm} \aleph(O_t, \mu_{jm}, U_{jm}) \tag{2.1}$$

où M est le nombre de gaussiennes utilisées pour modéliser la fonction de densité de probabilité dans chaque état, C_{jm}, μ_{jm} et U_{jm} sont respectivement le gain, le vecteur de moyennes et la matrice de covariances de la gaussienne m dans l'état j. Les gains des composantes gaussiennes C_{jm} satisfont les conditions suivantes :

$$\sum_{m=1}^{M} C_{jm} = 1 \qquad C_{jm} \geq 0 \qquad 1 \leq j \leq N, \qquad 1 \leq m \leq M$$

2.5.1.2 Les trois problèmes des MMCs

Il existe trois problèmes typiques que l'on peut chercher à résoudre avec un MMC [96] :

Problème de reconnaissance Etant donnés une séquence d'observations $O = (O_1, O_2, ..., O_T)$ et un modèle $\Lambda = (\pi, A, B)$, comment calculer la vraisemblance de la séquence, c'est-à-dire la probabilité que cette séquence d'observations ait été générée par ce modèle $P(O|\Lambda)$?

La probabilité d'une séquence d'observations se décompose sur toutes les séquences d'états S possibles dans un modèle Λ ayant émis la séquence d'observations :

$$P(O|\Lambda) = \sum_{q} P(O|q,\Lambda)\ P(q|\Lambda) = \sum_{q} \pi_{q_1} b_{q_1}(O_1)\, a_{q_1 q_2} b_{q_2}(O_2) ... a_{q_{T-1} q_T} b_{q_T}(O_T) \quad (2.2)$$

Cette équation nécessite des opérations d'ordre N^T, et donc n'est pas envisageable en pratique. Cette vraisemblance est alors calculée par programmation dynamique avec l'algorithme *Forward-Backward* ou par son approximation *Viterbi* [96].

Problème d'identification des états cachés Etant donnés une séquence d'observations et un modèle, comment déterminer la meilleure séquence d'états qui a donné naissance à la séquence d'observations ? Ce problème adresse la question de découvrir la séquence d'états cachés du modèle qui soit optimale selon un critère donné.

Le retour arrière (*backtracking*) après décodage par l'algorithme de *Viterbi* permet de trouver cette séquence d'états [96].

Problème d'apprentissage Etant donnée une séquence d'observations, comment ajuster les paramètres d'un modèle pour maximiser la probabilité de la séquence d'observation ?

L'apprentissage se fait selon l'algorithme de *Baum-Welch* [96], qui fait appel au principe de l'algorithme de EM, (Expectation Maximization [96])

Ces algorithmes de programmation dynamique rendent les MMC d'utilisation efficace et aisée. Le but de notre travail n'étant pas le développement théorique des MMCs, ces algorithmes ne sont pas détaillés dans cette section. Pour plus de détails concernant ces algorithmes et la mise en œuvre des Modèle de Markov Cachés, se référer au livre de *Rabiner* [96].

2.5.1.3 Modélisation de la signature par un MMC

Les Modèles de Markov Cachés (MMCs) sont bien adaptés à la signature manuscrite en-ligne. Dans la littérature, les MMCs sont utilisés pour modéliser la signature d'une personne à partir d'un ensemble de ses réalisations (signatures dédiées à l'apprentissage). Ces signatures sont représentées par une séquence de vecteurs, où chaque vecteur est une observation représentant un point (ou une portion) de la signature. Ainsi, chaque signature est une séquence d'observations, comme montré sur la Figure 2.9.

L'avantage des MMCs est qu'ils permettent, en modélisant stochastiquement les séquences d'observations de longueurs variables, de prendre en compte la variabilité des signatures et de traiter les distorsions non linéaires. De plus, ils incorporent le principe de programmation dynamique pour unifier la segmentation et la classification de séquence de signatures variant dans le temps.

Dans le cadre de la vérification des signatures en-ligne, les Modèles de Markov Cachés peuvent être discrets ou continus [96]. Plusieurs travaux de l'état de l'art privilégient l'utilisation des MMCs continus [65, 23, 24, 109, 110, 34]. En effet, comme les MMCs discrets [67, 101] considèrent que les observations proviennent d'un ensemble d'alphabets fini, dans beaucoup d'applications, une discrétisation des données continues est introduite, induisant par la suite une perte d'information et une dégradation du signal. Les MMCs continus tentent de pallier ce problème en considérant qu'une observation est générée suivant un mélange de gaussiennes [97] (voir Figure 2.9).

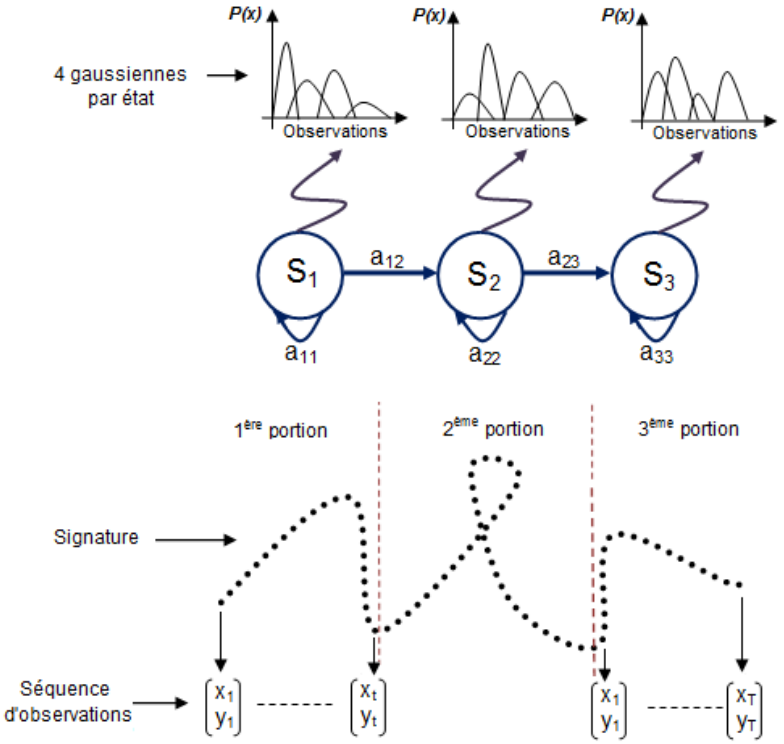

FIGURE 2.9: Modélisation de la signature par un Modèle de Markov Caché.

Le choix de la topologie (nombre d'états, nombre de gaussiennes dans chaque état, etc.) est crucial pour la modélisation des signatures par les MMCs. Une topologie gauche-droite est très souvent utilisée pour modéliser la signature manuscrite : elle n'autorise que les transitions d'un état vers lui-même ou vers l'état suivant. D'ailleurs, la signature manuscrite est un processus évolutif dans le temps et il n'existe pas de retour en arrière dans la signature. Cette organisation permet donc de modéliser des contraintes temporelles. Chaque état modélise une portion du signal de la signature, localement stationnaire, représentée dans l'espace des paramètres par un certain nombre de vecteurs alignés temporellement (voir Figure 2.9). Le MMC est alors bien approprié pour modéliser la signature en-ligne : les états décrivent les parties stationnaires dans la signature, alors que les probabilités de transitions décrivent les ruptures ou autrement dit les variations dans le signal de la signature en-ligne.

En outre, deux paramètres essentiels sont ordinairement considérés pour sélectionner la structure optimale du modèle MMC : le nombre d'états N et le nombre de gaussiennes par état M. Il n'y a pas de règle générale pour déterminer les valeurs optimales de ces deux paramètres. Cependant, certains travaux de la littérature privilégient l'utilisation d'un nombre d'états personnalisé [109, 110], dépendant de la longueur totale des signatures authentiques disponibles pour l'apprentissage du MMC. Quant au nombre de gaussiennes, la plupart des travaux de

la littérature utilisent un nombre de gaussiennes fixe, et un nombre de 4 gaussiennes par état est très souvent choisi [119, 23, 24, 109, 110].

Pour conclure cette partie, le processus général d'un système de vérification de signature en-ligne basé sur un Modèle de Markov Caché est décrit par le diagramme illustré sur la Figure 2.10. Un système de vérification de signature en-ligne basé sur un MMC comporte une phase d'apprentissage et une phase de vérification.

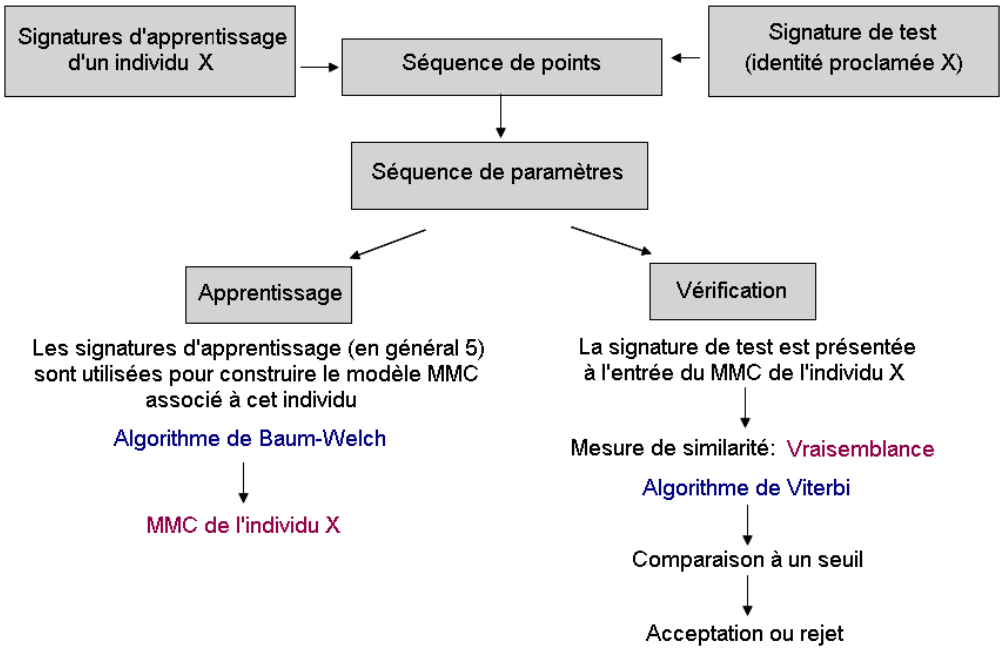

FIGURE 2.10: Processus général d'un système de vérification de signature en-ligne basé sur un Modèle de Markov Caché.

La phase d'apprentissage consiste en une estimation des paramètres de ce MMC obtenue en maximisant la vraisemblance des signatures d'apprentissage via l'algorithme Baum-Welch [96]. En phase de vérification, la vraisemblance entre la signature de test et le modèle (score de mise en correspondance), $P(O|\Lambda)$, est calculée avec l'algorithme de Viterbi [96], où O est la séquence des vecteurs d'observations multidimensionnels correspondant à la signature de test, Λ est le modèle associé à la personne dont les paramètres ont été appris lors de la phase d'apprentissage.

2.5.2 Distance élastique (DTW)

Etant donné que deux signatures n'ont pas exactement la même longueur, et que les décalages temporels ne sont pas linéaires, une comparaison directe entre deux signatures par le biais d'une distance n'est pas évidente. Afin de tenir compte

de cette variabilité, certains systèmes de vérification de la signature en-ligne reposent sur une distance dite "élastique" (*Dynamic Time Warping* ou *DTW*[96]).

La distance élastique est une technique de programmation dynamique dédiée à gérer les déformations temporelles. La distance élastique a été originellement proposée pour des applications de reconnaissance de la parole [37]. Elle permet de trouver pour chaque élément d'une séquence, le meilleur élément qui lui correspond dans l'autre séquence, relativement à un certain voisinage et à une certaine métrique(voir Figure 2.11).

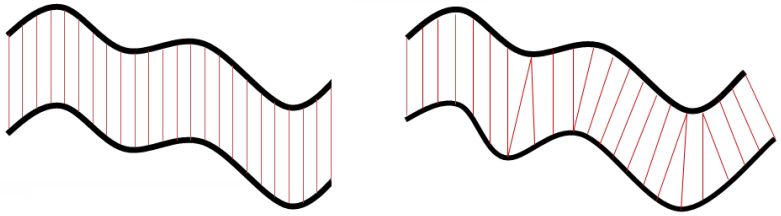

FIGURE 2.11: Mise en correspondance point à point entre deux signatures sans décalages temporels (à gauche), et avec décalages temporels en appliquant la distance élastique (à droite).

2.5.2.1 Algorithme de la distance élastique

La DTW est le coût cumulé minimal quand nous alignons de manière non linéaire deux séquences de différentes longueurs (voir Figure 2.11) pour quantifier la dissimilarité entre deux signatures.

Supposons qu'on ait une signature de référence (R) et une signature de test (T). Chacune d'elles est une séquence de points correspondant aux signaux acquis par le capteur : $R = r_1, ..., r_N$ et $T = t_1, ..., t_M$

L'algorithme DTW réalise d'abord un alignement non linéaire entre ces deux séquences, en construisant une matrice *(N*M)*, où chaque élément *(i,j)* contient la distance euclidienne entre deux points r_i et t_j. Puis, il recherche, parmi tous les alignements possibles, le chemin de déformation optimal qui minimise une fonction de coût cumulé. Cette dernière n'est autre que la distance cumulative $D(i,j)$ correspondant à la somme de toutes les distances euclidiennes locales rencontrées pour aller d'un point initial, correspondant au début des deux séquences, à un point final, correspondant à la fin des deux séquences :

$$D(i,j) = d(r_i, t_j) + min(D(p(i,j)))$$

Où *D(i,j)* est la distance globale au point *(i,j)*, $d(r_i, t_j)$ est la distance locale en ce point et *p(i,j)* est l'ensemble des prédécesseurs possibles de l'élément *(i,j)*.

La Figure 2.12 illustre le chemin optimal parcouru entre deux séquences de longueurs différentes.

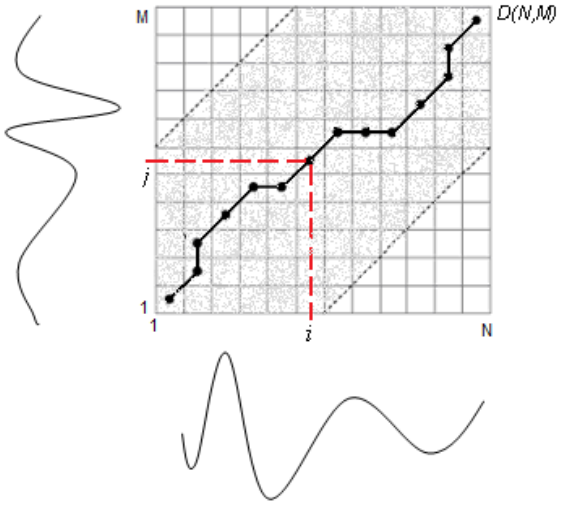

FIGURE 2.12: Chemin de déformation optimal.

Les prédécesseurs du point *(i,j)* sont choisis de manière à obtenir une trajectoire monotone et plausible. Un choix de prédécesseur possible et simple que nous utiliserons dans notre travail est celui illustré sur la Figure 2.13 : insertion, substitution et suppression.

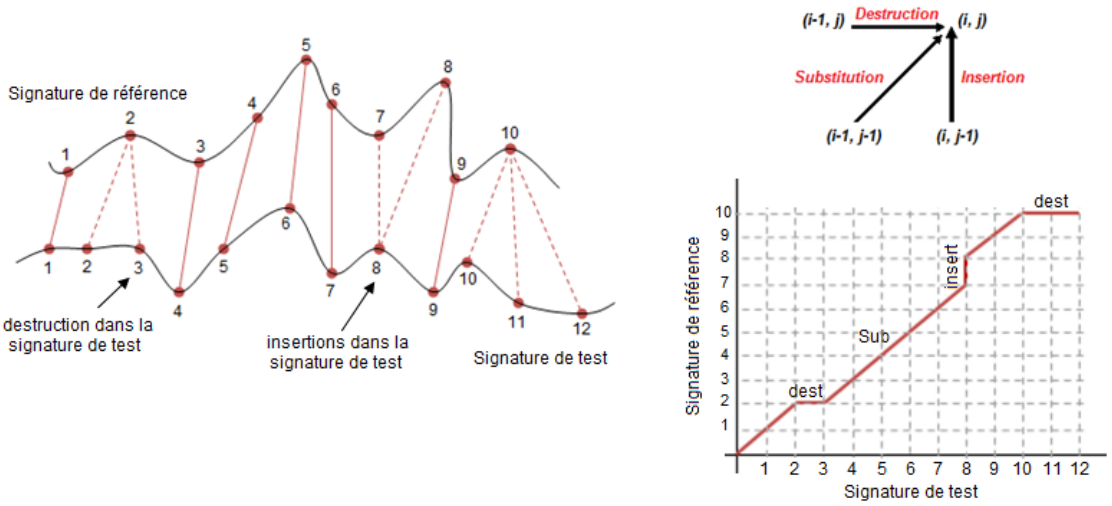

FIGURE 2.13: Type de déplacement : insertion, substitution et suppression.

L'algorithme avance alors sur le chemin qui minimise la distance :

$$D(i,j) = d(r_i, t_j) + min[D(i-1, j-1), D(i-1, j), D(i, j-1)]$$

Après avoir calculé la dernière distance *D(N,M)*, on obtient le tracé du chemin de déformation (voir Figure 2.12). La distance normalisée *D(N,M)/K* correspond à la distance entre R et T, où K est le nombre d'alignements.

En calculant les distances cumulatives, la matrice *(N*M)* est remplie de gauche à droite et de bas à haut. En conséquence, bien que la mise en œuvre reste très simple, la complexité du calcul est d'ordre *O(N*M)* et cela n'est pas acceptable en programmation. Pour alléger l'algorithme, on limite la région de recherche en utilisant la méthode de *Sakoe* [96], comme montré par la partie grisée sur la Figure 2.12 .

2.5.2.2 La distance élastique dans les systèmes de vérification

Le processus de vérification de signatures consiste à évaluer la distance DTW de la signature de test à chacune des références. La distance élastique est très utilisée dans le contexte de la vérification de signatures manuscrites en-ligne car elle est facile à mettre œuvre, et très bien adaptée pour une utilisation dans le cas où on a peu de signatures dans la base de référence.

Plusieurs variantes de la DTW ont par conséquent émergé dans le but d'apporter des améliorations à l'algorithme de base [62, 74, 72, 106]. En effet, certains travaux de l'état de l'art ont proposé de réduire le nombre de points représentatifs des signatures en conservant uniquement des points pertinents afin de diminuer le temps de calcul [114] ; d'autres ont privilégié d'effectuer la mise en correspondance entre des portions de signatures [17]. Par ailleurs, différentes définitions de la distance locale ont été adoptées : distance spatiale, temporelle ou curviligne [115].

2.5.3 Comparaison entre les deux approches MMC et DTW

Généralement, les approches reposant sur des modèles prennent plus de temps au moment de l'apprentissage que celles basées sur des références, mais elle sont plus rapides au moment du test. En effet, pour les MMCs, la signature de test est comparée à un seul modèle statistique par personne, qui modélise toutes les signatures d'une personne donnée, alors que pour la DTW, la signature de test est comparée à toutes les signatures disponibles de la personne dans la base de référence.

Bien que le fonctionnement de l'approche statistique par MMC diffère de celui de l'approche par DTW, le MMC et la DTW restent très concurrentes en terme de performances des systèmes de vérification. Néanmoins, pour arriver à obtenir de bonnes performances, c.à.d. un faible taux d'erreur égale *EER*, l'approche statistique (MMC) est très souvent associée à un plus grand nombre de paramètres locaux par rapport à l'approche à distance (DTW) [82]. Cela a été prouvé dans un travail effectué en interne au sein de notre laboratoire, mais qui n'a pas encore été publié. Nos expériences effectuées sur la base de signatures MCYT-100 ont montré qu'en utilisant seulement les coordonnées *(x,y)* comme paramètres, de bonnes performances peuvent être obtenues avec la distance élastique. Alors qu'avec le MMC, beaucoup plus de paramètres sont nécessaires pour atteindre les mêmes performances. Toutefois, si le MMC est associé à un codage (paramètres) efficace, il peut facilement surpasser la DTW en terme de performances.

Ce phénomène a été aussi observé lors de la compétition de vérification de signature en-ligne "SVC'2004" [122]. On distingue deux tâches dans cette compétition : dans la première tâche, seules les coordonnées du stylo sont disponibles ; dans la deuxième tâche, en plus des coordonnées, la pression et les deux angles d'inclinaison du stylo sont disponibles. Les résultats de SVC'2004 montrent que pour les deux tâches, le meilleur système est basé sur la distance élastique (DTW), proposé par *Kholmatov et al.* , décrit dans [71]. Ce système obtient un taux *EER* d'environ 2.8% sur les deux tâches. En deuxième place, on distingue les systèmes basés sur les MMCs, avec des taux *EER* d'au moins 4.6%. Nous avons constaté que les informations comme la pression et les angles d'inclinaison du stylo ne sont pas importantes pour le système basé sur la distance élastique : avec ou sans ces informations, ce système donne en effet presque les mêmes résultats (environ 2.8% *EER*). Par contre, sans ces informations, les taux *EER* des systèmes basés sur les MMCs augmentent, par exemple de 4.6% à 6.2% ou de 5.0% à 5.9%.

Si nous nous sommes attardés sur ce sujet, c'est parce que l'étude des performances dans les chapitres suivants sera effectuée avec les deux classifieurs MMC et DTW qui correspondent à l'état de l'art et cela en considérant uniquement les coordonnées *(x,y)* pour la représentation des signatures.

Comme nous le détaillerons dans le Chapitre 4, notre mesure de qualité de signature en-ligne est calculée localement en représentant cette dernière uniquement par les coordonnées *(x,y)*. Ceci nous permettra en premier temps d'analyser visuellement les résultats préliminaires, en se basant sur la forme de la signature. De plus, les coordonnées *(x,y)* sont les seuls paramètres existant dans toutes les

bases de signatures acquises sur une tablette graphique ou sur un PDA. Ainsi, l'étude de performances sera effectuée en ne considérant que *(x,y)*. L'utilisation des deux approches DTW et MMC n'est donc pas anodine : l'approche à distance (DTW) fonctionne très bien avec les coordonnées seulement ; et, comme nous voulions confronter notre mesure de qualité à différents systèmes de vérification, nous avons aussi utilisé une approche statistique basée sur les MMCs, qui sont couramment utilisés dans la littérature. Toutefois, comme cette thèse reprend tous nos travaux effectués en plusieurs parties, certaines fois nous avons utilisé la DTW et dans d'autres le MMC.

Nous rappelons que notre but n'est pas de comparer les approches de classification entre-elles, ni de les optimiser pour trouver le meilleur classifieur, mais juste de valider notre mesure de qualité. C'est pourquoi nous ne nous préoccuperons pas des performances médiocres qui seront obtenues avec le MMC, puisque comme nous venons de l'évoquer le choix des seuls paramètres *(x,y)* en entrée du MMC n'est pas optimal pour mesurer de bonne performance avec de modèle.

2.6 Conclusion

Dans ce chapitre, nous avons présenté dans un premier temps, les outils pour évaluer les performances de n'importe quel système de vérification de signatures en-ligne. Tous ces outils seront utilisés dans les chapitres suivants pour évaluer les performances des systèmes de vérification en fonction des mesures de qualité que nous allons proposer. L'étude des performances sera effectuée sur les différentes bases de signatures présentées dans ce chapitre, en utilisant les deux approches de classification décrites dans ce chapitre aussi, à savoir les Modèles de Markov Cachés et la distance élastique.

Chapitre 3

Etat de l'art : mesures de qualité des signatures dans la littérature

Depuis plus d'une vingtaine d'années, les travaux précurseurs menés dans le cadre de la signature manuscrite, portaient d'une part sur l'élaboration et la mise en œuvre de différentes approches de classification [23, 24, 34, 35, 36, 110, 92, 67, 87, 74, 101, 120, 72, 17, 40, 46, 114, 44, 13, 47, 99], et d'autre part sur la sélection des paramètres les plus discriminants [23, 34, 100, 113, 33, 80, 38, 81, 88, 61, 63, 104, 72]. Ces travaux avaient pour objectif commun l'amélioration des performances des systèmes de vérification automatiques. A nos jours, les recherches dans ce domaine ont montré leur limite dans la performance des systèmes de vérification.

Pour améliorer ces systèmes, d'autres travaux plus récents se sont intéressés à mesurer la qualité des signatures. Un lien est fait dans ces travaux entre les performances que l'on peut attendre du système de vérification et la qualité des signatures. En fait, plusieurs critères ont été considérés comme des mesures de qualité telles que la stabilité de la signature, sa complexité, et sa lisibilité, sans forcément qu'un lien soit fait entre ces différentes mesures.

Comme cette thèse porte sur la définition de nouvelles mesures de qualité, nous allons d'abord dresser un état des lieux de la littérature concernant les mesures de qualité dans le cadre de la signature manuscrite, qu'elle soit en-ligne ou hors-ligne, et nous montrerons l'impact de ces mesures sur les performances des systèmes de vérification.

3.1 Mesures de qualité des signatures authentiques

Comme déjà expliqué dans le Chapitre 2, un système de vérification de signature en-ligne requiert deux phases : une phase d'enregistrement et une phase de vérification. Pour un bon fonctionnement du système de vérification, il est primordial que les signatures authentiques destinées à l'enregistrement montrent une certaine stabilité et complexité : la stabilité est exigée pour bien caractériser l'individu par le biais de ses réalisations ; la complexité est exigée pour que la signature d'un individu soit difficilement reproductible par d'autres que lui. En effet, une signature peu complexe est facile à imiter, ce qui favorise l'augmentation du taux de fausses acceptations pour n'importe quel système de vérification. Une signature trop variable favorise non seulement l'augmentation du taux de faux rejets mais aussi le taux de fausses acceptations.

3.1.1 Mesures de qualité dans le contexte en-ligne

Dans la même optique avec ce qui vient d'être dit, *Brault et al.* proposent dans [14] d'accepter une signature d'une personne à l'enregistrement, seulement si elle est assez complexe et pas trop variable. Les auteurs ont donc proposé deux mesures différentes pour quantifier la complexité et la stabilité de la signature. *La complexité* est quantifiée par le biais d'un "coefficient de difficulté" qui estime la difficulté à reproduire chaque portion dans une signature donnée. Ce coefficient est basé sur le calcul d'un taux de modifications géométriques (longueur et direction des portions) par unité de temps. Par ailleurs, *la variabilité intra-classe* est quantifiée par le biais d'un "indice de dissimilarité", basé sur le calcul de la distance élastique (DTW) entre toutes les signatures d'enregistrement de l'individu. De ce travail, les auteurs ont conclu que les signataires "problématiques" en termes de performance des systèmes de vérification sont ceux dont les signatures présentent une grande variabilité intra-classe et un faible "coefficient de difficulté", c.à.d. une signature peu complexe. Nous voyons alors apparaître dès 1989, pour la première fois les notions de variabilité et complexité d'une signature manuscrite.

Dans [21], *Dimauro et al.* ont proposé une technique pour évaluer *la stabilité* locale des signatures en-ligne afin d'améliorer la procédure de vérification des systèmes automatiques. Basé sur le calcul de la distance élastique (DTW) entre toutes les signatures authentiques d'un individu, le critère de stabilité associé à

cet individu correspond au nombre de points participant à une mise en correspondance directe, c.à.d. seule la substitution entre deux signatures intra-classe est prise en compte. Plus ce nombre est grand, plus la signature est considérée comme stable. Les auteurs ont ensuite incorporé cette information de stabilité à la phase de classification. Leur approche repose sur le postulat suivant : plus les portions de la signature sont stables, plus elles sont difficiles à imiter, car une plus grande précision est exigée pour imiter des portions générées de manière assez semblable par le signataire. Ainsi, le degré de stabilité des portions de la signature a été utilisé pour estimer la pertinence de la décision de vérification de ces portions. Les résultats ont montré que les portions de la signature les plus stables sont les plus difficiles à imiter, et que l'information de stabilité peut être utile pour l'amélioration des performances des systèmes de vérification.

En utilisant cette même technique basée sur *la stabilité* locale des signatures en-ligne, un autre travail [76] réalisé par la même équipe a été effectué dans le but de sélectionner les meilleures signatures de référence. Les signatures sélectionnées comme références sont celles contenant le plus de portions stables. Les résultats ont montré l'efficacité de cette technique à réduire le taux de fausses acceptations dans les systèmes de vérification automatiques.

Dans [77], *Liu et al.* ont aussi utilisé la notion de *stabilité* pour sélectionner les signatures de référence. La stabilité a été quantifiée en utilisant la distance élastique (DTW) : pour chaque signature authentique d'une personne, la valeur moyenne des distances DTW ont été calculées entre cette signature et toutes les autres signatures authentiques de la personne. Plus cette distance moyenne est faible, plus la signature authentique est considérée similaire aux autres et donc sélectionnée pour la base de référence.

Dans [85], *Müler et al.* ont introduit le terme "qualité" dans leur travail. Ils se sont attaqués au problème de l'évaluation de la qualité des signatures en-ligne en utilisant deux méthodes différentes. Dans la 1$^{\text{ère}}$ méthode, la qualité des signatures en-ligne dépend d'un classifieur et est associée au taux d'erreur *EER* du système de vérification. Cette méthode repose sur le postulat suivant : la mesure de qualité d'une signature authentique est une expression quantitative qui prédit son utilité dans la séparation entre les deux classes, à savoir la classe des authentiques et celle des imitations. En utilisant un classifieur basé sur la distance élastique (DTW) évalué sur la base MCYT-100, la qualité d'une signature authentique est mesurée par l'*EER*, qui est calculé en comparant cette dernière à toutes les autres signatures authentiques ainsi qu'aux imitations. Plus l'*EER* associé à cette

signature authentique est faible, plus cette signature est considérée comme étant une bonne candidate contribuant à une meilleure séparation entre les signatures authentiques et les imitations.

Dans la 2ème méthode, la qualité des signatures en-ligne est liée à *la stabilité intra-classe* des signatures authentiques calculée comme dans [77]. La stabilité a été considérée comme étant la valeur moyenne des distances DTW calculées entre la signature authentique considérée et toutes les autres signatures authentiques. Plus cette distance moyenne est faible, plus la signature authentique considérée est similaire aux autres, et plus elle est jugée stable. En utilisant un système de vérification basé sur une distance élastique (DTW), les expériences effectuées sur la base MCYT-100 ont montré que la stabilité des signatures de référence a un impact crucial sur les performances de ce système. Par ailleurs, les auteurs ont aussi étudié l'influence de différents paramètres sur la stabilité des signatures. Ils ont trouvé que certains paramètres, comme la vitesse moyenne et le nombre de portions, ont un faible impact sur la stabilité des signatures.

Toujours dans le contexte en-ligne, un autre travail est à citer, celui de *Richiardi et al.* [98]. Ce travail est différent des autres évoqués précédemment, du fait que la mesure de qualité proposée dépend entièrement d'un modèle probabiliste basé sur la distance de Mahalanobis [108]. En effet, cette mesure de qualité est basée sur les propriétés de la matrice de covariance de chaque composante gaussienne dans les Modèles à Mélanges de Gaussiennes (GMM)[97]. La mesure de qualité utilisée correspond au déterminant de la matrice de covariance : plus le déterminant est proche de zéro, plus le résultat de la distance de Malalanobis sera biaisé, et donc la classification sera faussée. Après évaluation d'un système de vérification basé un Modèle à Mélange de Gaussiennes (GMM) [97] sur la base MCYT-100, les résultats ont montré une forte corrélation entre la mesure de qualité proposée et les scores résultant de la classification. Cette mesure de qualité basée sur la matrice de covariance est donc une manière efficace pour prédire les performances du système basé sur un GMM.

3.1.2 Mesures de qualité dans le contexte hors-ligne

Dans le cadre de la signature hors-ligne, plusieurs travaux ont été menés dans le but d'étudier la qualité des signatures authentiques, afin de prédire les performances des systèmes de vérification en tenant compte principalement de trois critères : la stabilité de la signature, sa complexité et sa lisibilité.

Dans [2], *Allgrove et al.* ont évalué la qualité de la base de référence en s'appuyant sur le critère de *stabilité* des signatures de référence. Dans ce travail, la stabilité est quantifiée en se basant sur la minimisation de la distance euclidienne calculée entre tous les vecteurs de paramètres globaux extraits des signatures de la base de référence. Les résultats ont montré que cette procédure de validation des signatures de référence contribue à l'amélioration des performances d'un système de vérification.

Dans [3, 32], *Alonso-Fernandez et al.* ont proposé des critères quantitatifs de *complexité* et de *variabilité*. La mesure de *complexité* est basée sur les directions d'inclinaison existant dans la signature. En effet, certaines signatures présentent beaucoup d'intersections avec différentes inclinaisons, comme montré sur la Figure 3.1.b. Pour ces signatures, il n'existe donc pas une direction d'inclinaison prédominante, à l'inverse des autres signatures qui ne contiennent pas plusieurs intersections, comme montré sur la Figure 3.1.a. Ainsi, la complexité d'une signature a été quantifiée comme étant l'aire de cette signature contenant des directions d'inclinaison non prédominantes. Puis, à chaque personne lui est associée une mesure de complexité moyennée sur toutes les signatures de référence. Quant à la mesure de *variabilité intra-classe*, elle est basée sur le calcul de la distance de Mahalanobis [108] entre chaque signature de référence et le modèle statistique. En moyennant toutes ces distances liées aux signatures de référence, à chaque personne lui est associée une mesure de variabilité intra-classe.

FIGURE 3.1: Exemples de signatures de MCYT-75 présentant des directions d'inclinaion a) prédominantes b) non prédominantes.

Afin d'analyser l'impact de ces deux mesures sur les performances, deux classifieurs ont été évalués sur la base de signature hors-ligne MCYT-75 : un classifieur global basé sur la distance de Mahalanobis [108], et un classifieur local basé sur les Modèles de Markov Cachés [96]. Après avoir ordonné les signataires de la base en fonction des deux mesures de complexité et de variabilité séparément, les expériences ont montré que les performances se dégradent pour les personnes dont leur signature est très variable, et pour les personnes dont leur signature est peu complexe, c.à.d. à faible aire de directions d'inclinaison non prédominantes, comme la signature illustrée sur la Figure 3.1.a.

Fairhurst et al. dans [31] ont eux aussi exploité les critères de *variabilité* et *complexité*. Cependant, dans ce travail, ces deux critères ont été évalués visuellement par des experts humains. Un opérateur humain étiquette d'abord les signatures selon les deux critères, puis il classifie les signatures comme étant authentiques ou imitations pour mesurer l'impact de chacun des deux critères sur les performances de classification. Afin d'analyser l'impact de la variabilité sur les erreurs produites lors de la vérification par les experts humains, les auteurs ont utilisé le coefficient de corrélation Spearman. Les résultats ont montré qu'il existe une forte corrélation positive entre la variabilité et le taux de faux rejets (FRR), et une corrélation négative entre la variabilité et le taux de fausses acceptations (FAR). De ce fait, les auteurs ont conclu que plus la variabilité intra-classe est grande, plus le FRR est grand, et plus le FAR est faible. D'autre part, pour étudier l'impact de la complexité de la signature sur les performances de vérification sans être affecté par la variabilité, deux cas de figure ont été pris en compte : lorsque les signatures sont stables, et lorsque les signatures sont instables. Pour les signatures stables, les résultats ont montré que plus la complexité augmente, plus le FRR augmente et plus le FAR diminue. Dans le cas contraire où les signatures sont instables, une augmentation de la complexité semble causer une confusion chez les experts humains entraînant simultanément une augmentation des FRR et des FAR.

Dans le cadre de la signature hors-ligne, un autre critère a été introduit pour caractériser les signatures des individus, à savoir le critère de *lisibilité*. Dans [32, 4], *la lisibilité* et *la complexité* ont été évalués visuellement par des experts humains. Les signatures de la base hors-ligne MCYT-75 ont été classées manuellement en 3 catégories suivant le critère de lisibilité, et en 4 catégories suivant le critère de complexité. Afin d'étudier l'influence de ces critères sur les performances, deux classifieurs ont été utilisés pour la phase de vérification : un classifieur global basé sur une distance de Mahalanobis [108], et un classifieur local basé sur les Modèles de Markov Cachés [96]. En considérant les différentes catégories obtenues séparément, les meilleures performances ont été obtenues pour les signatures les plus complexes et les plus lisibles, c.à.d. les signature cursives, celles qui se rapprochent le plus de l'écriture.

Dans [11, 12], *Boulétreau et al.* ont proposé une nouvelle mesure de qualité, totalement différente de ce que nous avons vu jusqu'à présent, basée sur la dimension fractale. Cette mesure permet de catégoriser les signataires automatiquement suivant le critère de *lisibilité*. Des méthodes de classification automatiques par agrégation ont été exploitées pour afficher un "graphe de lisibilité" sur lequel les classes obtenues ont été illustrées et caractérisées par différents styles d'écriture,

comme montré sur la Figure 3.2 : signatures "hautement cursives", signatures "très cursives", signatures "avec écritures séparées", signatures "avec écritures mal formées", et signatures "paraphes simples". Cependant, aucune évaluation de performance n'a été menée sur ces catégories.

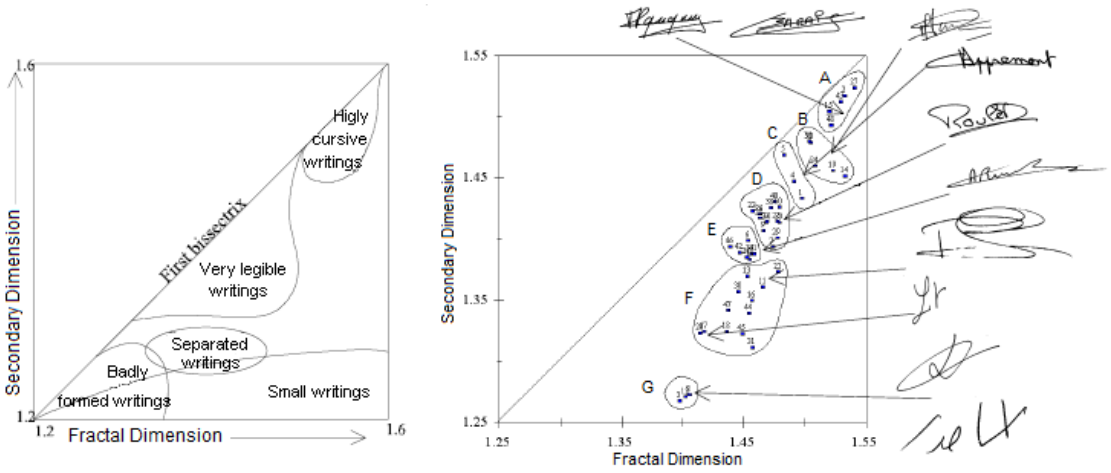

FIGURE 3.2: Graphe de lisibilité.

3.2 Mesures de qualité étendues aux "imitations"

Jusqu'à présent, tous les travaux de la l'état de l'art que nous avons évoqué se sont penchés sur l'étude des mesures de qualité des signatures authentiques uniquement. Cependant, la fiabilité d'un système de vérification de signature dépend de sa capacité à discriminer les signatures authentiques des imitations. Cette capacité est liée à deux facteurs : premièrement, à la bonne caractérisation d'une personne par le biais de ses réalisations, et deuxièmement, à la non-vulnérabilité des signatures aux attaques.

Ces deux facteurs sont reflétés par les deux types d'erreurs générés par les systèmes biométriques, les faux rejets et les fausses acceptations. De ce fait, les performances des systèmes de vérification dépendent non seulement de la qualité des signatures authentiques d'un individu, comme vu plus haut dans la Section 3.1 de ce troisième chapitre, mais aussi de leur vulnérabilité aux attaques. La plupart des travaux de la littérature concernant les mesures de qualité se sont focalisés sur la qualité des signatures authentiques, alors que très peu ont évalué la difficulté à imiter une signature authentique. Dans cette section, nous présentons les deux principaux travaux de la littérature dédiés à l'étude de la vulnérabilité des signatures authentiques aux attaques.

Dans le cadre de la signature en-ligne, à notre connaissance, la seule étude effectuée dans ce sens est celle menée par *Brault et al.* [15]. Le but de ce travail est d'estimer quantitativement et a priori à partir des coordonnées de la signature seulement, la difficulté à reproduire l'image de la signature ainsi que le geste de signer. Pour cela, un "coefficient de difficulté" lié à la complexité du tracé de la signature authentique a été calculé, de la même façon que celui décrit dans la Section 3.1.1 de ce chapitre. Afin de vérifier l'efficacité de ce coefficient, les auteurs ont utilisé deux méthodes de comparaison : une méthode subjective, basée sur l'opinion des imitateurs ; et une méthode objective, basée sur le calcul d'une dissimilarité spatio-temporelle entre chaque signature authentique et toutes ses imitations. Les auteurs ont conclu que la difficulté à imiter une signature est liée aux variations locales de l'aspect spatial de la signature au niveau des portions : plus les variations locales sont fortes, et plus la signature est difficile à imiter.

Dans le cadre de la vérification de l'écriture manuscrite en-ligne, une étude intéressante a été effectuée en vue de son application dans le domaine de la sécurité [6]. *Ballard et al.* ont tout d'abord généré empiriquement 3 catégories de styles d'écriture : écritures dites en "blocs" présentant beaucoup de levés de stylos entre les lettres (voir Figure 3.3.a) ; écritures dites "cursives" (voir Figure 3.3.c), pour lesquelles la plupart des lettres sont attachées ; et des écritures dites "mixtes", pour lesquelles quelques lettres sont attachées (voir Figure 3.3.b). Par la suite, considérant 4 types d'imitations de différentes qualités (imitations aléatoires, statiques, dynamiques, et imitations entraînées faites par de bons imposteurs qui ont imité seulement les écritures ressemblant aux leurs), les auteurs ont conclu que les écritures "en blocs" et "mixtes" sont les plus faciles à imiter par rapport aux écritures "cursives".

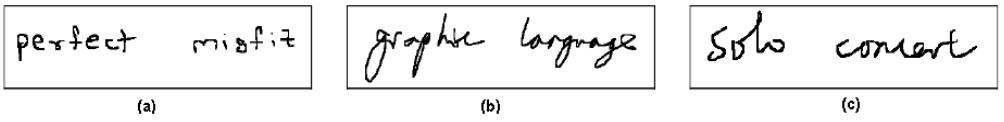

FIGURE 3.3: Exemple d'écritures (a) "en blocs", (b) "mixtes" et (c) "cursives".

3.3 Catégories de personnes

Jusqu'à présent nous avons présenté les différents critères de la littérature utilisés pour caractériser la qualité de la signature manuscrite. Nous avons noté dans certains travaux cités dans ce chapitre que ces critères ont induit naturellement la formation de catégories de personnes générées soit visuellement par des

experts humains [32, 4, 6], soit en ordonnant les personnes selon les valeurs du critère considéré [3, 32, 11, 12].

En analysant ces catégories de personnes, les différents auteurs sont d'accord sur le fait que le degré de difficulté de reconnaissance diffère d'une personne à une autre ; de plus, il existe une catégorie de personnes dont les signatures sont plus fiables en termes de performances.

Dans ce même esprit, des travaux initiés par *Doddington* dans le cadre des modalités visage, parole et empreintes digitales ont proposé de catégoriser les personnes en fonction de leur comportement vis-à-vis d'un système de reconnaissance automatique [22, 116, 117, 10, 9]. Cette catégorisation est connue sous le nom de "Ménagerie Biométrique" (ou *Biometric Menagerie*) car des noms d'animaux ont été octroyés aux différentes catégories de personnes. Concernant la modalité parole, les personnes ont été classées en 4 catégories [22] : Moutons (ou *Sheeps*, locuteurs faciles à reconnaître), Chèvres (ou *Goats*, locuteurs difficiles à reconnaître), Agneaux (ou *Lambs*, locuteurs faciles à imiter) et Loups (ou *Wolves*, les bon imitateurs). Ces catégories ont été définies en se basant soit sur la moyenne des scores des données authentiques des personnes, soit sur la moyenne des scores des imitations.

Cette classification a été appliquée plus tard à la modalité visage [116] et complétée par une autre classification : Vers (ou *Worms*, les pires personnes en termes de performances), Caméléons (ou *Chameleons*), Fantômes (ou *Phantoms*) et Colombes (ou *Doves*, les meilleures personnes en termes de performances). Et cela, en se basant cette fois-ci sur les scores des données authentiques ainsi que sur les scores des imitations simultanément [117, 118]. Cette dernière catégorisation a été ensuite appliquée à la modalités empreintes digitales [117, 118]. La Figure 3.4 illustre les différentes catégories de personnes de la "Ménagerie Biométrique" générées suivant les scores des données authentiques et ceux des imitations.

A noter que ces catégories ne sont pas nécessairement toutes présentes dans les différents systèmes de reconnaissance, car elles dépendent du classifieur utilisé et de la qualité des données disponibles dans la base de données. Par ailleurs, cette catégorisation est très connue des chercheurs travaillant sur les modalités visage, empreintes digitales et parole, mais n'a jamais été étendue à la modalité signature manuscrite.

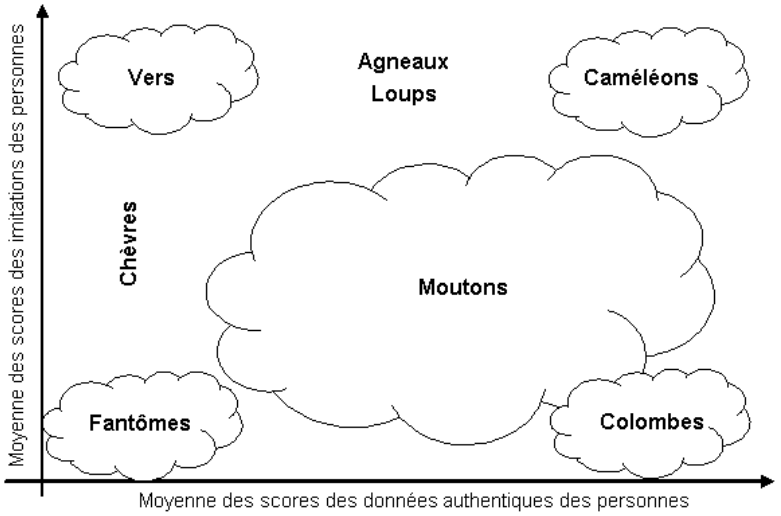

FIGURE 3.4: Catégories de personnes de la "Ménagerie Biométrique" générées en fonction des scores des données authentiques et des scores des imitations.

3.4 Conclusion

Dans ce chapitre, nous avons présenté les travaux de l'état de l'art qui concernent les mesures de qualité appliquées à la signature manuscrite. Nous avons remarqué que dans la littérature émergent plusieurs critères de qualité qui semblent "naturels" : complexité, variabilité et lisibilité. Ceux-ci sont calculés de manière plus ou moins similaire dans les différents travaux et un lien est fait avec les performances des systèmes de vérification correspondants. Nous notons cependant qu'il faut plusieurs mesures pour bien caractériser un système. Par ailleurs, nous avons noté le manque de travaux abordant le problème de mesurer la qualité des imitations.

Nous verrons dans les chapitres suivants que notre travail de thèse est extrêmement novateur car nous proposons d'abord dans le Chapitre 4 une mesure de qualité des signatures authentiques qui peut englober les 3 critères usuels de la littérature : la complexité, la stabilité et la lisibilité [42, 49, 43]. Puis, nous proposons dans le Chapitre 6 une autre mesure qui englobe simultanément ces 3 critères usuels ainsi que la vulnérabilité des signatures aux attaques. Nous verrons aussi que ces deux nouvelles mesure permettent de catégoriser les personnes et d'identifier celles qui sont les plus fiables en termes de performances des systèmes de vérification ainsi que les moins fiables, et ceci indépendamment des systèmes de vérification considérés.

Chapitre 4

Mesure de qualité des signatures authentiques basée sur l'entropie

Dans ce chapitre, nous allons introduire notre nouvelle mesure de qualité des signatures en-ligne basée sur le concept d'entropie. Il est à noter que seule la qualité des signatures authentiques est traitée dans ce chapitre.

Dans ce qui suit, nous commencerons par présenter la notion d'entropie, concept fondamental dans la théorie de l'information reposant sur le calcul probabiliste. Nous ferons ensuite un bref état de l'art concernant l'utilisation de l'entropie comme mesure de qualité en biométrie. Puis, nous décrirons deux manières de calculer notre mesure de qualité à partir de la notion d'entropie, soit à partir d'une estimation globale des densités de probabilité par le biais des Modèles à Mélange de Gaussiennes , soit à partir d'une estimation locale des densités de probabilité par le biais des Modèles de Markov Cachés. Une fois les deux procédures de calcul expliquées, nous nous intéresserons à la possibilité de générer des catégories de personnes suivant la qualité de leur signature mesurée par notre critère d'entropie. Ceci nous conduira à favoriser la deuxième approche à base de Modèles de Markov Cachés. Nous essayerons ensuite de faire un lien entre notre mesure d'entropie et les critères de qualité proposés dans la littérature. Enfin, nous confronterons notre mesure de qualité aux performances de deux systèmes de vérification classiques dans la mesure où nous évaluerons les performances des systèmes sur chaque catégorie de personnes générée avec l'entropie et ferons un lien entre dégradation de qualité des signatures et perte de performances.

4.1 Entropie dans la théorie de l'information

4.1.1 Introduction à la théorie de l'information

La théorie de l'information ne s'est érigée en discipline autonome que très progressivement, renforcée au fil du temps par les notions et concepts mathématiques et physiques qu'elle a engendrés [18]. Les premiers travaux visant la mesure de l'information datent des années vingt, mais ce n'est qu'à partir de 1948, grâce aux travaux de *Shannon* [18] que la théorie de l'information a pris sa forme actuelle.

La théorie de l'information est une discipline fondamentale qui s'applique généralement dans le domaine de la communication [102]. Elle correspond principalement à trois problématiques et tente de répondre aux questions suivantes :

1. *Mesurer l'information* : disposant d'une source de données,

   - Quelle est la quantité d'information ?

   - Quelle est la complexité de ces données ?

2. *Comprimer l'information* :

   - Quel est le taux de compression maximal sans perte d'information ?

3. *Transmettre l'information* : dans le cas où les données traversent un système où règnent des perturbations,

   - Quel est l'effet des perturbations sur l'information ?

   - Quel est le taux de transmission maximal sans perte d'information ?

Notre travail concerne uniquement le premier point et s'attache à mesurer la quantité d'information apportée par une source d'information, qui dans notre cas est la signature authentique d'une personne donnée. Cette quantité sera mesurée par le biais de l'entropie [18], concept fondamental de la théorie de l'information, qui repose essentiellement sur la théorie des probabilités. Nous allons donc maintenant présenter ce concept introduit par *Shannon* dans les années 50.

4.1.2 Entropie de Shannon

L'entropie est une quantité qui désigne en général la part de désordre, de dégradation ou de hasard que comporte tout système d'information [18].

La théorie de l'information est basée sur une description probabiliste des données et des systèmes, qui sont modélisés à l'aide de variables aléatoires. L'entropie d'une variable aléatoire est une mesure quantitative de l'incertitude (ou, alternativement, de la quantité d'information) associée aux valeurs prises par la variable aléatoire [18].

Considérons pour une variable aléatoire discrète X, un évènement $X{=}x$ de probabilité $p(x)$. Intuitivement, plus $p(x)$ est faible, plus x est improbable et plus sa réalisation sera "inattendue". On dit alors que l'évènement $X{=}x$ apporte "beaucoup d'information" [18]. Au contraire, si $p(x)$ est grand, la réalisation de $X{=}x$ est très probable et n'est pas incertaine et apporte donc "peu d'information" [18] ; elle est même quasi certaine si $p(x) \simeq 1$.

Il apparaît donc qu'il existe un lien entre l'information fournie par une source et la distribution de probabilité de la sortie de cette source [18]. Pour mesurer l'information apportée par la réalisation $X{=}x$, ou en d'autres termes le manque d'information (ou l'incertitude) sur $X{=}x$, on calcule la quantité :

$$- log\ p(x) \tag{4.1}$$

C'est en effet une quantité positive, qui est très grande quand $p(x)$ est petit, et est très faible quand $p(x)$ est grand (proche de 1).

L'entropie d'une variable aléatoire discrète X est définie comme une information moyenne sur toutes les valeurs possibles de x :

$$H(X) = -\sum_{x} p(x)\ log\ p(x) \tag{4.2}$$

L'entropie mesure le manque d'information ; elle nous donne alors la quantité totale d'information qui fait défaut. Aussi, il faut noter que l'entropie de la variable aléatoire X dépend uniquement de la distribution des probabilités $p(x)$, et pas des valeurs x appartenant à X prises par la variable elle-même [18].

4.2 L'entropie mesure de qualité en Biométrie

Plusieurs travaux de la littérature relatifs au domaine de la biométrie ont déjà fait usage du concept d'entropie. Basé sur un calcul probabiliste, l'entropie

a été appliquée à plusieurs modalités en adoptant des démarches différentes pour l'estimation des probabilités.

En effet, dans le cadre de la modalité iris, l'entropie a été utilisée par *Daugman* d'une part pour mesurer le contenu d'information dans l'image [19], et d'autre part pour mesurer l'unicité/la singularité d'un iris [20]. Dans [19], *Daugman* a mesuré l'entropie d'une image de 8 bits-pixels en niveaux de gris, où les probabilités correspondent à la fréquence de chaque niveau de gris dans l'image. Dans un travail plus récent [20], *Daugman* a proposé le concept d'entropie discriminante pour quantifier la corrélation entre les bits de modèles iris (degrés de liberté). *Daugman* a constaté que pour un modèle iris à 2048 bits, l'incertitude ou le désordre concerne 249 bits non corrélés. La correspondance des bits de deux codes d'iris de longueur m revient à appliquer m tests de Bernoulli dont la probabilité de succès (deux bits alignés sont les mêmes) est de 0,5. Ainsi, la distance de Hamming entre les codes d'iris est une variable aléatoire dont la fonction de densité de probabilité a une distribution binomiale. L'entropie de discrimination des codes d'iris est alors modélisée en utilisant la moyenne et l'écart-type de la distribution binomiale des distances de Hamming entre les codes d'iris.

Toujours pour un but de discrimination, *Adler et al.* [1] et *Jassim et al.* [64] ont proposé dans le cadre de la modalité visage une mesure d'entropie relative par personne qui correspond à la quantité d'information qui distingue la personne d'une population donnée, en considérant à chaque fois différents paramètres. L'entropie a été calculée suivant une loi continue multidimensionnelle en utilisant une distribution gaussienne des densités de probabilité, où la moyenne et l'écart-type ont été calculés indépendamment pour chaque paramètre sur un ensemble d'images.

Dans le cadre de la signature manuscrite [70], le concept d'entropie a aussi été utilisé pour mesurer l'unicité d'une signature, autrement dit combien il est probable que deux personnes aient accidentellement la même signature. Pour cela, *Kholmatov et al.* ont calculé d'abord les coefficients de transformée de fourier à partir de la seule coordonnée y de la signature. L'entropie a été calculée en se basant sur la probabilité que k coefficients de Fourier parmi les n de la signature test correspondent à ceux de la signature référence. Cette probabilité suit une loi binomiale. Pour trouver le nombre d'harmoniques n requises pour représenter la signature, les auteurs se sont basés sur la minimisation du taux d'erreur *EER* du classifieur. Puis, pour n fixé, ils ont trouvé le nombre k de coefficients qui correspondent entre la signature test de référence à l'*EER*. Ainsi, les auteurs ont

trouvé que l'entropie d'une signature garantissant son unicité est entre 11 et 19 bits.

L'entropie a également été utilisée comme mesure de qualité pour améliorer la robustesse des classifeurs biométriques à la dégradation de la qualité des données [7]. En fait, *Bendris et al.* ont introduit l'entropie comme mesure de qualité dans le processus de classification d'un système biométrique multimodal (audiovisuel) en utilisant les deux informations audio et image. Les auteurs ont combiné les scores des 2 classifieurs (audio et image) en utilisant les mesures de qualité de l'image et de l'audio. La fusion est basée sur une somme pondérée des scores des deux systèmes : le score audio a été pondéré par le ratio de l'énergie moyenne de la séquence du signal audio et celle du signal bruit ; le score de l'image a été pondéré par la mesure d'entropie basée sur la probabilité qu'un pixel choisi de l'image ait une valeur d'intensité entre 1 et 256 (256 gradients utilisés pour encoder l'image).

Par ailleurs, le concept d'entropie a été utilisé dans le cadre de l'analyse des documents [66]. *Jung-Tae et al.* ont proposé une mesure de complexité qui reflète le degré de difficulté de reconnaissance d'un caractère image dans une base de données. Cette mesure est définie par le calcul de l'entropie de chaque position *(x,y)* de l'image dans la base, en utilisant la probabilité qu'un point noir dans cette position apparaisse à la même position dans les différentes images de la base. Puis, la mesure de complexité est obtenue en additionnant toutes les entropies de toutes les positions dans l'image.

D'après les travaux de la littérature cités dans cette section, on constate que le concept d'entropie n'a été utilisé dans le cadre de la signature manuscrite que dans un seul travail, celui effectué par *kholmatov et al.* [70]. Dans ce travail, l'entropie mesure l'individualité de la signature, c.à.d. combien il est probable que deux personnes aient accidentellement la même signature. Et cela, en se basant sur un classifieur. Alors que notre travail a pour but d'introduire une mesure de qualité indépendante de toute étape de classification, qui mesure la quantité d'information dans la signature d'une personne en se basant sur plusieurs instances de cette signature. Dans ce qui suit, nous présenterons les deux approches que nous avons adoptées pour le calcul de la mesure d'entropie.

4.3 Calcul de la mesure d'entropie

L'entropie mesure donc l'incertitude liée à une variable aléatoire basée sur sa distribution [18]. Notre objectif étant de mesurer la quantité d'information contenue dans une signature donnée, deux questions se sont posées alors au tout début de notre étude :

1. Comment modéliser une signature en-ligne par une variable aléatoire ?

2. Comment estimer la distribution de cette variable aléatoire ?

Pour répondre à ces deux questions, nous avons dans un premier temps, fait le choix de représenter les signatures uniquement par les coordonnées *(x,y)*. Ce choix est motivé d'une part par la possibilité d'analyser alors visuellement les résultats en se basant sur l'aspect spatial de la signature, et d'autre part par le fait que les coordonnées *(x,y)* sont les seuls paramètres existant dans toutes les bases de signatures, qu'elles soient acquises sur une tablette graphique ou sur un PDA.

Ensuite, nous avons considéré deux cas de figure : dans le premier cas, nous avons proposé d'estimer les densités de probabilité globalement via un Modèle à Mélange de Gaussiennes (GMM) [97], et dans le deuxième cas localement via un Modèle de Markov Caché (MMC) [96].

4.3.1 Mesure d'entropie avec un GMM

Comme nous voulons mesurer la qualité d'une signature donnée, il nous a paru tout à fait naturel d'estimer la densité de probabilité globalement sur toute cette signature. Pour cela, nous faisons intervenir les Modèles à Mélange de Gaussiennes (*GMM, Gaussian Mixture Models*) [97] qui ne sont autres que des Modèles de Markov Cachés [96] à un seul état et M gaussiennes.

Ainsi, nous considérons chaque point dans la signature, autrement dit chaque observation *(x,y)*, comme étant une réalisation d'une variable aléatoire z=*(x,y)* qui suit une loi de probabilité discrète, où z appartient à l'alphabet Z de couples *(x,y)* appartenant à toute la signature (voir Figure 4.1).

L'entropie $H_m(Z)$ associée à chaque gaussienne m, mesurée en bits, est alors calculée comme suit :

$$H_m(Z) = -\sum_{z \in Z} p(z)\ log_2\ p(z) \tag{4.3}$$

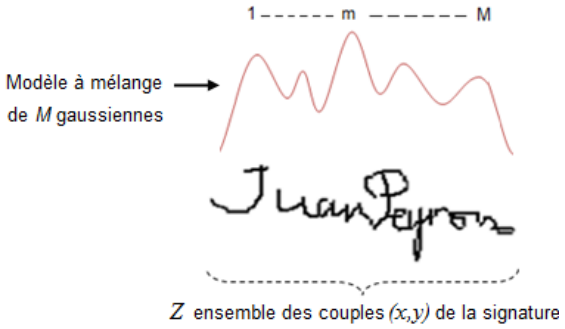

FIGURE 4.1: Modélisation de la signature par un Modèle à Mélange de Gaussiennes.

où Z correspond à l'ensemble des points appartenant à la signature.

L'entropie de chaque signature $H_{sig}(Z)$ est obtenue en moyennant sur toutes les gaussiennes de cette dernière, la valeur de l'entropie associée à chaque gaussienne. Puis, une normalisation par la durée de cette signature T est effectuée afin de comparer les signatures de différentes personnes entre-elles. En effet, l'entropie d'une signature est une somme de probabilités estimées localement sur tous les points de la signature. De ce fait, plus la signature est longue, plus l'entropie a tendance à être grande et vice versa.

L'entropie d'une signature $H_{sig}(Z)$ est alors définie comme suit :

$$H_{sig}(Z) = \frac{1}{T * M} \sum_{m=1}^{M} H(Z_m) \tag{4.4}$$

où M est le nombre de gaussiennes de la signature, T est la durée de la signature en secondes calculée comme suit :

$$T = \frac{Nombre\,de\,points\,de\,la\,signature}{Fréquence\,d'échantillonnage\,du\,capteur} \tag{4.5}$$

Cette mesure d'entropie ainsi calculée donne le nombre de bits par seconde nécessaires en moyenne pour décrire la quantité d'information contenue dans toute la signature considérée.

La Figure 4.2 illustre le processus de calcul de l'entropie d'une signature donnée par le biais d'un GMM.

Pour calculer les probabilités discrètes, nous exploitons la loi continue estimée sur toute la signature par le Modèle à Mélange de Gaussiennes (GMM) associé à la personne considérée.

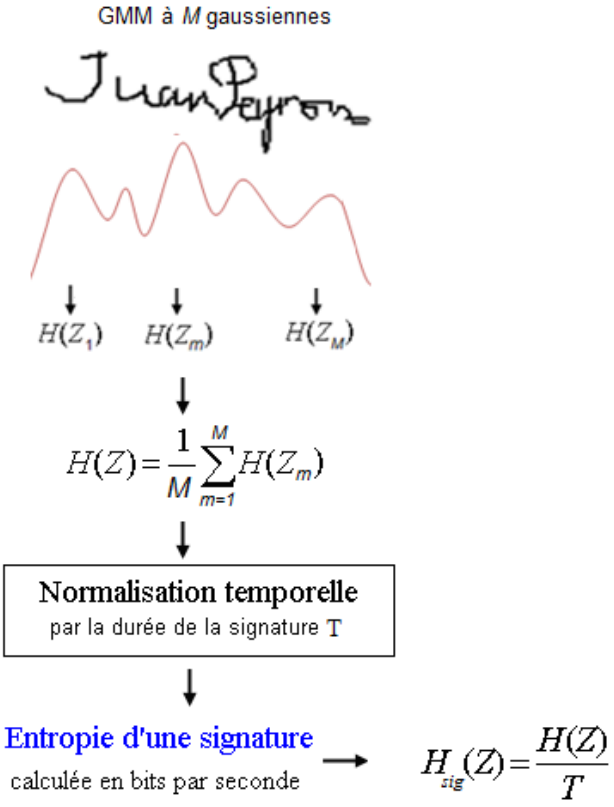

FIGURE 4.2: Calcul de l'entropie d'une signature avec un Modèle à Mélange de Gaussiennes.

En effet, la signature en-ligne est un signal temporel considéré comme un processus continu à partir duquel on récupère une séquence de valeurs discrètes via le capteur (tablette graphique ou PDA). Pour cette raison, bien que *z=(x,y)* soit une variable aléatoire discrète, nous profitons de la loi de probabilité continue estimée sur toute la signature par le GMM de la personne. La fonction de densité, dans le cas du GMM, est modélisée par un mélange de lois gaussiennes.

De cette manière, pour calculer l'entropie associée à la signature d'une personne donnée, nous entraînons le GMM de la personne sur K signatures authentiques de ce scripteur, après calcul d'un nombre de gaussiennes personnalisé M, comme suit :

$$M = \frac{T_{total}}{30} \tag{4.6}$$

où T_{total} est le nombre total de points disponibles dans la base d'apprentissage du scripteur, et nous considérons empiriquement qu'il faut au moins 30 points pour bien estimer le vecteur de moyenne et la matrice de covariance d'une gaussienne.

Par la suite, nous exploitons le GMM de la personne et considérons les densités de probabilités estimées par le GMM pour calculer l'entropie de la signature

globalement suivant l'équation 4.4.

Enfin, l'entropie associée à la personne, c.à.d. son Entropie Personnelle H_p, est calculée en moyennant les valeurs d'entropie $H_{sig}(Z)$ des K signatures, sur lesquelles les densités de probabilité ont été estimées par le GMM :

$$H_p = \frac{1}{K} \sum_{sig=1}^{K} H_{sig}(Z) \tag{4.7}$$

L'Entropie Personnelle est mesurée en bits par seconde. Les expériences ont montré qu'il est nécessaire d'avoir au minimum K=10 signatures par personne afin d'obtenir une bonne estimation de l'Entropie Personnelle : en effet, au-delà de 10 signatures dans la base d'apprentissage, on a remarqué que la valeur d'Entropie Personnelle reste relativement stable.

4.3.2 Mesure d'entropie avec un MMC

Dans le deuxième cas de figure, nous avons proposé de calculer l'entropie en estimant les densités de probabilité localement, sur des portions de la signature [43]. Cette proposition repose sur le postulat que la signature en-ligne est un signal temporel caractérisé par des variations locales, entre lesquelles le signal reste relativement stable.

Etant donné que le signal de la signature en-ligne est stationnaire par morceaux, nous avons fait appel dans ce cas aux Modèles de Markov Cachés [96].

De cette manière, nous considérons chaque signature comme une suite de portions, générées par la segmentation obtenue via un Modèle de Markov Caché (MMC) de la personne. En effet, le MMC est un outil naturel pour modéliser les signatures, car :

- il génère des portions de façon automatique par minimisation d'un critère global sur la signature (Segmentation par algorithme de Viterbi [96]).

- il estime localement la densité de probabilité multi-gaussienne sur chaque portion par l'algorithme Estimation-Maximisation [96].

Ainsi, nous obtenons autant de portions dans chaque signature que d'états dans le MMC de la personne appris sur des réalisations de sa signature. Nous considérons alors chaque point *(x,y)* dans une portion donnée S_i comme étant une

réalisation d'une variable aléatoire $z=(x,y)$ qui suit une loi de probabilité discrète, où z appartient à l'alphabet Z_i de couples (x,y) appartenant à la portion S_i (voir la Figure 4.3). L'entropie en chaque portion S_i, mesurée en bits, est alors calculée comme suit :

$$H(Z_i) = -\sum_{z \in Z_i} p(z)\ log_2\ p(z) \tag{4.8}$$

Cette quantité donne le nombre de bits nécessaires en moyenne pour décrire le contenu d'information en chaque portion S_i de la signature considérée [43].

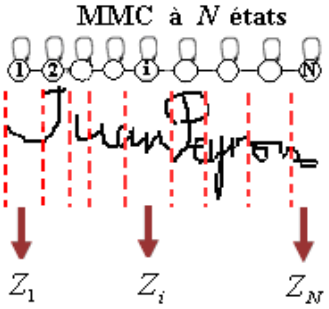

FIGURE 4.3: Modélisation de la signature avec un Modèle de Markov Caché.

Comme pour le GMM, afin de calculer les probabilités discrètes, nous exploitons la loi continue estimée en chaque portion par le Modèle de Markov Caché associé à la personne considérée [43]. Chaque fonction de densité est modélisée par un mélange de lois gaussiennes.

Pour calculer l'entropie associée à la signature d'une personne donnée [43], nous entraînons le MMC de la personne sur K signatures authentiques de ce scripteur, après calcul d'un nombre d'états personnalisé N, comme suit :

$$N = \frac{T_{total}}{M * 30} \tag{4.9}$$

où T_{total} est le nombre total de points disponibles dans la base d'apprentissage du scripteur, et M est le nombre de composantes gaussiennes par état. En se basant sur les travaux de la littérature, nous avons fixé $M = 4$. Nous considérons aussi qu'il faut au minimum 30 points par état pour bien estimer le vecteur moyenne et la matrice de covariance d'une gaussienne.

Par la suite, nous exploitons le MMC de la personne pour segmenter chacune de ses K signatures par l'algorithme de Viterbi [96]. Puis, pour chaque portion, nous considérons les densités de probabilités estimées par le MMC pour calculer l'entropie localement suivant l'équation 4.8.

La Figure 4.4 illustre le processus de calcul de l'entropie d'une signature donnée.

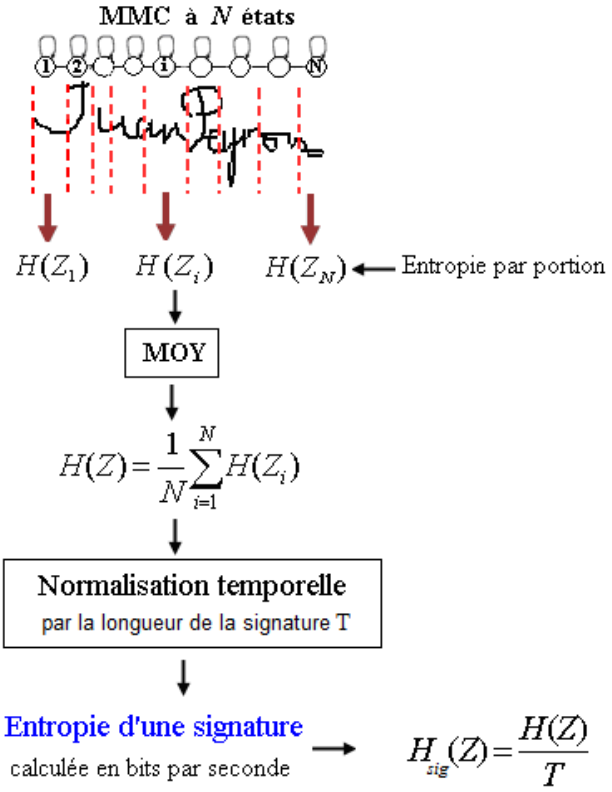

FIGURE 4.4: Calcul de l'entropie d'une signature avec un Modèle de Markov Caché.

L'entropie de chaque signature $H_{sig}(Z)$ est obtenue en moyennant sur toutes les portions de cette dernière, la valeur de l'entropie associée à chaque portion (voir la Figure 4.4). Puis, une normalisation par la durée de cette signature T est effectuée afin de comparer les signatures de différentes personnes entre-elles [43]. L'entropie d'une signature $H_{sig}(Z)$ est alors définie comme suit :

$$H_{sig}(Z) = \frac{1}{T * N} \sum_{i=1}^{N} H(Z_i) \tag{4.10}$$

où N est le nombre de portions de la signature (égal au nombre d'états du MMC entraîné sur les K signatures authentiques), et T est la durée de la signature en secondes calculée comme précédemment dans la Section 4.3.1, suivant l'équation 4.5.

Enfin, l'entropie associée à la personne, c.à.d. son Entropie Personnelle H_p, est calculée en moyennant les valeurs d'entropie $H_{sig}(Z)$ des K signatures, sur lesquelles les densités de probabilité ont été estimées par le MMC :

$$H_p = \frac{1}{K} \sum_{sig=1}^{K} H_{sig}(Z) \tag{4.11}$$

L'Entropie Personnelle est mesurée en bits par seconde [43]. Pour les mêmes raisons que dans la Section 4.3.1, l'Entropie Personnelle est calculée en considérant K=10 signatures pour le MMC [43].

4.4 Catégories de personnes générées par la mesure d'Entropie Personnelle

Dans cette partie, les expériences sont effectuées sur deux bases de signatures, à savoir la base MCYT-100 [91] et le sous-ensemble BioSecure DS2-104 [54, 90]. Ces bases ont été décrites en détail dans le Chapitre 2. Pour chaque personne, l'Entropie Personnelle est calculée suivant les deux approches décrites dans la section précédente, et ce pour les 100 personnes de la base MCYT et les 104 personnes de la base BioSecure DS2.

Par la suite, nous nous sommes intéressés à la possibilité de former des catégories de personnes suivant leurs valeurs d'Entropie Personnelle et nous avons, pour cela, effectué une classification automatique non supervisée. Nous avons utilisé une classification hiérarchique ascendante [89] en considérant toutes les valeurs d'Entropie Personnelle de la base considérée. Puis, nous avons procédé à l'étude du nombre de catégories optimal en faisant appel à différents indices de validité [45]. La procédure suivie pour cette étude est détaillée en Annexe A.

Nous avons réalisé cette étude pour les deux cas de figure, c.à.d. pour les valeurs d'Entropie Personnelle calculées globalement avec un GMM et celles calculées localement avec un MMC. Les résultats ont montré l'existence d'une séparation optimale des signatures de chacune des bases (MCYT-100 et DS2-104) en 3 catégories de personnes : catégorie à "Haute Entropie", catégorie à "Moyenne Entropie" et catégorie à "Basse Entropie".

Cependant, nous avons constaté que la séparation entre les différentes catégories de personnes est beaucoup plus nette pour les valeurs d'Entropie Personnelle calculées localement avec le MMC. Le Tableau 4.1 montre la différence des valeurs d'Entropie calculées avec un GMM et celles calculées avec un MMC sur les 3 catégories d'entropie générées.

Statistiques	GMM			MMC		
	Haute Entropie	Moyenne Entropie	Basse Entropie	Haute Entropie	Moyenne Entropie	Basse Entropie
Pourcentage de personnes	5%	46%	49%	5%	28%	67%
Valeur moyenne d'Entropie	1,25	0,32	0,13	8,41	2,55	0,79
Ecart-type d'Entropie	0,25	0,11	0,03	0,98	1,00	0,34

TABLEAU 4.1: Distribution des personnes de MCYT-100 en chaque catégorie d'entropie

En effet, les valeurs d'Entropie Personnelle calculées globalement avec un GMM sont assez proches entre-elles, plus particulièrement pour les personnes de la catégorie "Moyenne" et "Basse Entropie", comme montré dans le Tableau 4.1. Ceci est expliqué par le fait que le GMM a tendance à lisser les densités de probabilité, et il est peu sensible aux variations locales dans la signature que le MMC.

Pour cette raison, dans le reste de notre travail de thèse, nous nous concentrons uniquement sur une représentation locale de la signature, et notre mesure de qualité sera calculée en se basant uniquement sur une estimation locale des densités de probabilité en chaque portion de la signature, via les Modèles de Markov Cachés.

La Figure 4.5 montre la distribution des personnes des bases MCYT-100 et DS2-104 dans chaque catégorie d'Entropie Personnelle calculée localement avec un MMC.

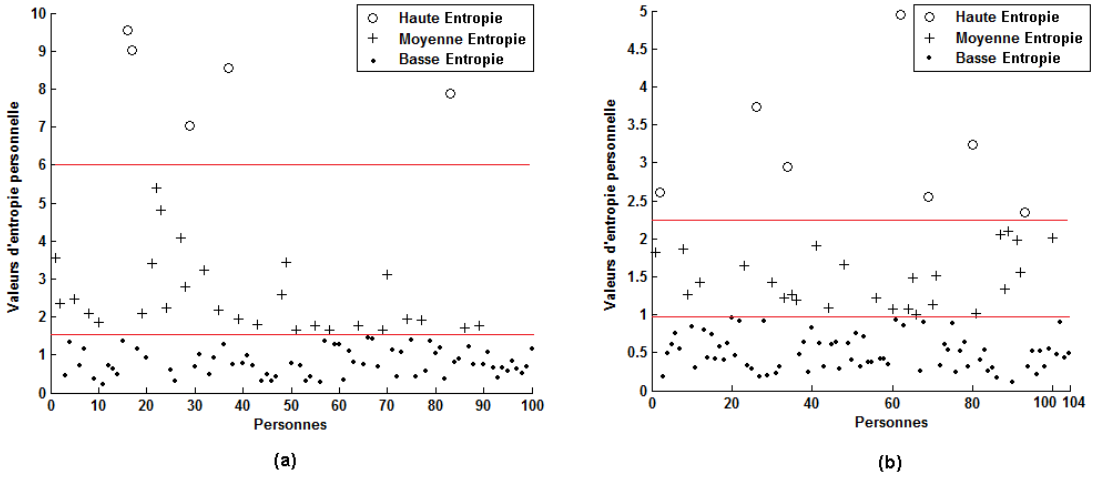

FIGURE 4.5: Valeurs d'Entropie Personnelle pour les personnes de (a) MCYT-100 et (b) DS2-104 par catégorie de personnes : (o) Haute, (+) Moyenne et (.) Basse Entropie Personnelle.

On constate que les 3 catégories obtenues sont effectivement linéairement séparables, comme représenté par les deux lignes horizontales. En comparant les

deux graphes de la Figure 4.5, on constate que le domaine de variation de l'Entropie Personnelle est différent sur MCYT-100 et DS2-104, et les valeurs d'Entropie sont à peu près 2 fois plus grandes dans MCYT-100 (Figure 4.5.a) que dans DS2-104 (Figure 4.5.b). En effet, les valeurs d'Entropie Personnelle dans MCYT-100 appartiennent à l'intervalle [0,233 ; 9,561], alors que dans DS2-104 elles appartiennent à l'intervalle [0,106 ; 4,946].

Les Tableaux 4.2 et 4.3 donnent le nombre de personnes par catégorie dans MCYT-100 et DS2-104, ainsi que la valeur moyenne et l'écart-type d'Entropie en chaque catégorie.

Statistiques	Catégories de personnes de MCYT-100		
	Haute Entropie	Moyenne Entropie	Basse Entropie
Pourcentage de personnes	5%	28%	67%
Valeur moyenne d'Entropie	8,41	2,55	0,79
Écart-type d'Entropie	0,98	1,00	0,34

TABLEAU 4.2: Distribution des personnes de MCYT-100 dans chaque catégorie d'Entropie Personnelle.

Statistiques	Catégories de personnes de DS2-104		
	Haute Entropie	Moyenne Entropie	Basse Entropie
Pourcentage de personnes	6,73%	25%	68,27%
Valeur moyenne d'Entropie	3,19	1,47	0,49
Écart-type d'Entropie	0,83	0,35	0,22

TABLEAU 4.3: Distribution des personnes de DS2-104 dans chaque catégorie d'Entropie Personnelle.

Ces tableaux montrent bien que la distribution des personnes en chaque catégorie est similaire pour les deux bases. En effet, entre 67% et 68% des personnes dans les deux bases ont des signatures à "Basse Entropie", tandis que les personnes ayant des signatures à "Haute Entropie" ne constituent que 5% à 6% de la base.

En outre, en observant les valeurs moyennes d'Entropie, on remarque qu'il existe un facteur 3 entre les catégories "Moyenne" et "Basse Entropie" pour les deux bases. De plus, ce facteur augmente entre les catégories extrêmes ("Haute" et "Basse Entropie") et atteint la valeur de 10 pour MCYT-100 et 6,5 pour DS2-104. Quant à la dispersion de chaque catégorie, la catégorie "Basse Entropie" est la plus compacte pour les deux bases, comme montré dans les Tableaux 4.2 et 4.3.

Afin d'analyser les résultats obtenus avec notre nouvelle mesure d'Entropie Personnelle, nous nous sommes tout d'abord basés sur l'aspect visuel des signatures. Ceci est possible car comme nous l'avons indiqué précédemment, l'Entropie

Personnelle est calculée en considérant les coordonnées *(x,y)* seulement, autrement dit, l'Entropie Personnelle dans ces expérimentations ne tient compte que de l'aspect spatial des signatures.

La Figure 4.6 et la Figure 4.7 montrent des exemples de signatures de chaque catégorie, respectivement sur MCYT-100 et DS2-104. A noter que les signatures de MCYT-100 qui seront montrées dans ce manuscrit ont toutes été déjà publiées dans [32, 3, 4, 91]. Quant aux signatures de DS2-104, leurs propriétaires ont autorisé leur publication.

FIGURE 4.6: Signatures de MCYT-100 à (a) Haute, (b) Moyenne et (c) Basse Entropie Personnelle.

FIGURE 4.7: Signatures de DS2-104 à (a) Haute, (b) Moyenne et (c) Basse Entropie Personnelle.

On remarque visuellement que les 3 catégories sont cohérentes. En effet, la première catégorie de personnes, à "Haute Entropie", contient des signatures courtes, simples et instables, qui ont souvent l'aspect d'un paraphe. A l'opposé, les signatures de la troisième catégorie, à "Basse Entropie", sont les plus longues, les plus complexes et les plus stables, se rapprochant parfois de l'écriture cursive. Entre ces deux catégories, les signatures à "Moyenne Entropie" (deuxième catégorie) correspondent à une catégorie de transition entre ces deux extrêmes, et sont plus longues que celles de la première catégorie et ont souvent l'aspect de paraphes complexes.

A noter que dans le cas où on aurait considéré les catégories de personnes générées avec l'Entropie Personnelle calculée globalement avec un GMM, la distinction entre les catégories aurait été moins nette. En effet, la Figure 4.8 montre des exemples de signatures paraphes (Figure 4.8.a) qui ont été classées dans la même catégorie ("Basse Entropie") que les signatures cursives (Figure 4.8.b) en se basant sur les valeurs d'Entropie calculées avec le GMM.

FIGURE 4.8: Exemples de signatures réparties dans la catégorie "Basse Entropie" suivant leur valeur d'Entropie calculée globalement avec un GMM.

Finalement, les 3 catégories de signatures obtenues avec la mesure d'Entropie Personnelle locale semblent à ce stade visuellement liées à la complexité et à la variabilité des signatures. Pour quantifier ces observations, nous proposons dans la section suivante, deux mesures quantitatives de complexité et de variabilité, avec lesquelles nous allons analyser plus finement les catégories d'Entropie obtenues.

4.5 Lien entre Entropie Personnelle, complexité et variabilité

4.5.1 Mesure de complexité

Afin de mesurer la complexité d'une signature donnée *"sig"*, nous avons considéré un vecteur V_{sig} à 7 composantes liées à la géométrie de la signature [43] : le nombre d'extrema locaux dans les deux directions x et y, les changements de direction du stylo dans les deux directions x et y, le nombre de points présentant des intersections simples (voir Figure 4.9.a), des intersections multiples (voir Figure 4.9.b) et des rebroussements (voir Figure 4.9.c) [16]. Puis, comme mesure de complexité de la signature, nous avons calculé la norme euclidienne de ce vecteur $\|V_{sig}\|$.

Par la suite, nous avons moyenné cette mesure sur les 10 signatures authentiques qui ont servi au calcul de l'Entropie Personnelle, afin de produire une mesure

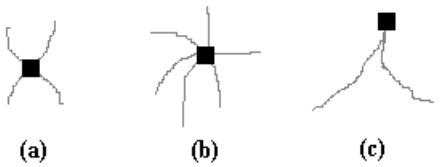

FIGURE 4.9: Illustration de points (a) d'intersection simple, (b) d'intersections multiples et (c) de rebroussement.

de complexité par personne C_p [43], ce qui permet de la confronter ensuite à notre mesure d'Entropie Personnelle :

$$C_p = \frac{1}{10} \sum_{sig=1}^{10} \|V_{sig}\| \tag{4.12}$$

où *sig* correspond à la signature authentique considérée.

4.5.2 Mesure de variabilité intra-classe

Afin de mesurer la variabilité intra-classe d'une personne, nous avons choisi d'utiliser la distance élastique (DTW) [96], qui repose sur un paradigme local pour quantifier les distorsions.

Dans ce cas, la signature de la personne a été représentée par 4 caractéristiques extraites localement en chaque point : la vitesse absolue, l'angle entre le vecteur vitesse absolue et l'axe horizontal, le rayon de courbure, et le rapport longueur / largeur sur une fenêtre glissante de taille 5. Puis, nous avons calculé les distances DTW deux à deux entre tous les couples possibles des 10 signatures authentiques considérées précédemment dans le calcul d'Entropie Personnelle (45 distances sont donc obtenues). Par la suite, la moyenne de ces 45 distances, V_p, a été considérée comme mesure de variabilité intra-classe de la personne considérée [43] :

$$V_p = \frac{1}{\binom{2}{10}} \sum_{\substack{sig,\, sig'=1 \\ sig \neq sig'}}^{10} DTW(s_{sig}, s_{sig'}) \tag{4.13}$$

où, *sig* et *sig′* sont deux signatures authentiques de la même personne.

4.5.3 Entropie Personnelle vs. complexité et variabilité

Les Figures 4.10 et 4.11 montrent les valeurs d'Entropie Personnelle en fonction des mesures de complexité et de variabilité par catégorie de personnes, respectivement sur MCYT-100 et DS2-104.

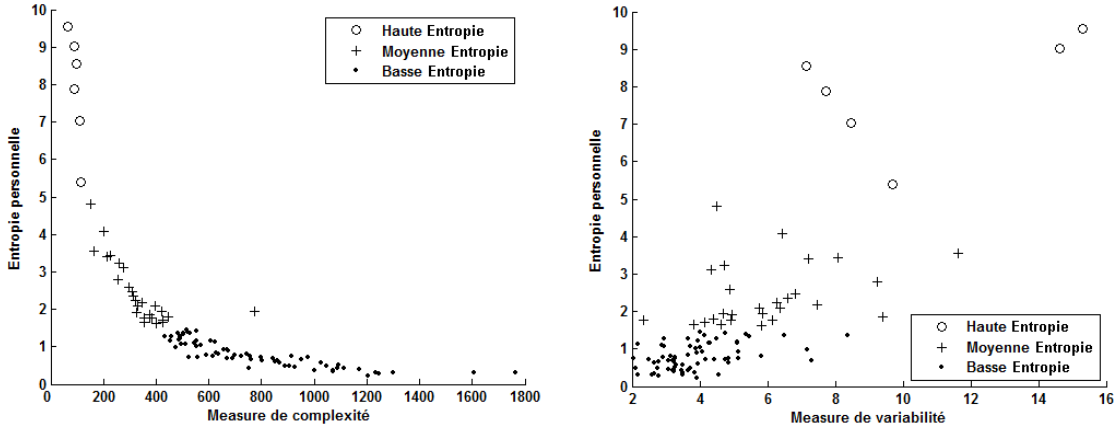

FIGURE 4.10: Entropie Personnelle vs. complexité (à gauche) et Entropie Personnelle vs. variabilité (à droite) par catégorie de personnes de MCYT-100.

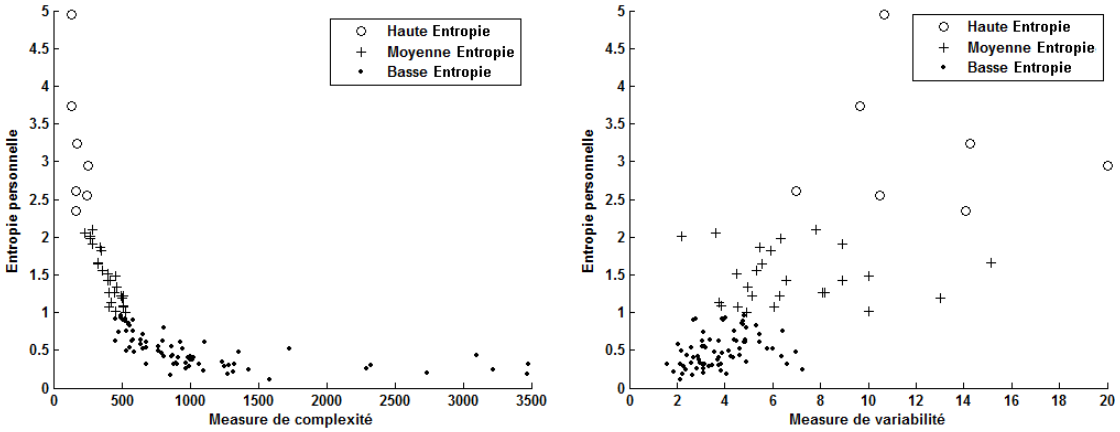

FIGURE 4.11: Entropie Personnelle vs. complexité (à gauche) et Entropie Personnelle vs. variabilité (à droite) par catégorie de personnes de DS2-104.

Les résultats obtenus confortent nos observations visuelles de la section précédente : les Figures 4.10 et 4.11 montrent bien que les signatures à "Haute Entropie", désignées sur les graphiques par 'o', sont très variables et peu complexes.

A l'opposé, les signatures à "Basse Entropie", indiquées par '.', sont de loin les plus complexes et les plus stables (montrant une faible variabilité). Tandis que les signatures à "Moyenne Entropie", indiquées par '+', correspondent à une catégorie de transition entre ces deux catégories extrêmes en termes des deux critères de complexité et de variabilité.

Ce comportement est vérifié pour les deux bases de signatures. Nous pouvons donc conclure que notre mesure d'Entropie Personnelle permet de quantifier simultanément la complexité de la signature ainsi que sa variabilité [43].

Ce résultat est très intéressant et renforce l'intérêt porté à notre nouvelle mesure de qualité basée sur l'entropie. En effet, comme évoqué dans le Chapitre 2, plusieurs travaux de la littérature ont considéré la complexité et la variabilité comme des critères de qualité des signatures. Cependant, dans certains travaux, ces critères ont été évalués uniquement visuellement par des experts humains ; dans d'autres, ils ont été quantifiés puis utilisés séparément. En comparaison à l'état de l'art, nous pouvons alors prétendre que notre nouvelle mesure de qualité a l'avantage de proposer une seule mesure qui englobe les critères de complexité et variabilité habituels. Ceci s'explique d'une part par le fait que l'entropie permet par définition de quantifier la complexité des données ; et d'autre part, le fait de calculer l'entropie d'une personne en estimant les probabilités sur un ensemble de réalisations du scripteur à l'aide des MMCs, permet de tenir compte de la variabilité intra-classe de ce scripteur.

Pour aller plus loin dans notre analyse, nous nous intéresserons, dans la partie qui suit, à l'évaluation des performances des systèmes de vérification sur les 3 catégories de personnes générées avec l'Entropie Personnelle. Comme ces catégories présentent des caractéristiques très différentes, les questions qui se posent à ce stade consistent à savoir comment un système de vérification automatique va se comporter en étant confronté à ces 3 catégories ? Y a-t-il une catégorie de personnes pour laquelle le système de vérification sera plus fiable en terme de performances ? Le comportement de ces catégories reste-t-il stable pour différents systèmes de vérification ? La section suivante sera dédiée à apporter des réponses à toutes ces questions.

4.6 Evaluation des performances des systèmes par catégorie de personnes

Cette section est consacrée à l'étude du comportement des 3 catégories de personnes générées avec l'Entropie Personnelle vis-à-vis des performances de différents systèmes de vérification.

Les expériences sont effectuées sur les deux bases MCYT-100 [91] et Bio-Secure DS2-104 [54, 90], en considérant les deux types d'imitations, à savoir les

bonnes imitations et les imitations aléatoires. Comme nous l'avons dit précédemment dans le Chapitre 2, nous avons fait le choix d'exploiter les deux classifieurs les plus utilisés dans la littérature : l'un basé sur la distance élastique (DTW) [96] et l'autre sur les Modèles de Markov Cachés (MMCs) [96].

4.6.1 Description des classifieurs et protocole d'évaluation

Comme pour le calcul de l'Entropie Personnelle, l'étude des performances est effectuée sur des signatures représentées uniquement par les coordonnées *(x,y)*.

Pour l'évaluation des performances, le protocole d'expérimentation que nous avons adopté est comme suit : 20 tirages aléatoires sont réalisés sur les signatures authentiques et les imitations. Chaque tirage contient 5 signatures authentiques utilisées comme signatures de référence. Pour le test, les 20 signatures authentiques restantes et les 25 bonnes imitations sont utilisées dans le cas de MCYT-100. Pour DS2-104, les tests sont effectués sur les 25 signatures authentiques restantes et les 20 bonnes imitations (appartenant aux 2 sessions). De plus, pour chaque personne 20 imitations aléatoires, correspondant aux signatures authentiques des autres personnes, sont considérées pour les deux bases. Ces imitations aléatoires sont choisies en nombre égal suivant les deux styles d'écriture des signatures : paraphes et cursives, à partir des catégories extrêmes. Les taux de fausses acceptations et de faux rejets sont calculés sur l'ensemble des 20 tirages aléatoires.

Pour le classifieur à distance élastique, le score de dissimilarité est défini comme étant le minimum des 5 distances DTW calculées entre la signature de test et les 5 signatures de référence :

$$Score_{DTW} = D_{min}(test, reference_i), \;\; i = 1...5 \tag{4.14}$$

En ce qui concerne la topologie du classifieur statistique, nous avons utilisé un MMC gauche-droite, avec un nombre d'états fixe pour toutes les personnes. Afin de garantir un compromis entre les signatures des deux catégories extrêmes, nous avons fait le choix de considérer un MMC à 6 états et 4 composantes gaussiennes par état. Le score de dissimilarité correspondant au classifieur MMC est le suivant :

$$Score_{HMM} = |LL - LL_{BA}| \tag{4.15}$$

où LL est la log-vraisemblance de la signature test (normalisée par la longueur de cette dernière), et LL_{BA} est la log-vraisemblance moyenne de la base d'apprentissage.

Il est important de remarquer que le MMC utilisé comme classifieur diffère de celui utilisé pour le calcul de l'Entropie. Le premier est consacré à la classification et est appris sur 5 signatures authentiques de la personne. Tandis que le second n'est exploité que pour l'estimation des densités de probabilité locales et il est appris sur 10 signatures authentiques de la personne. Ainsi, le nombre d'états par personne ne sera pas le même avec les deux MMCs puisque, d'après l'équation 4.9, le nombre d'état dépend de la longueur totale des signatures d'apprentissage. Par conséquent, les densités de probabilité estimées sur les signatures authentiques d'une même personne seront différentes avec ces deux Modèles de Markov Cachés.

Il faut noter aussi que les deux classifieurs DTW et MMC ont été intentionnellement utilisés sans aucune optimisation pour ces expériences. En effet, notre but n'est pas d'améliorer les performances des systèmes ou de comparer les différentes approches entre-elles mais d'analyser la différence relative dans les performances entre chaque catégorie de personnes.

4.6.2 Expérimentations et résultats

Les Figures 4.12 et 4.13 correspondant à MCYT-100 et les Figures 4.14 et 4.15 correspondant à DS2-104 montrent les courbes de performances par catégorie de personnes avec les deux classifieurs MMC et DTW respectivement, et avec les deux types d'imitations (bonnes et aléatoires).

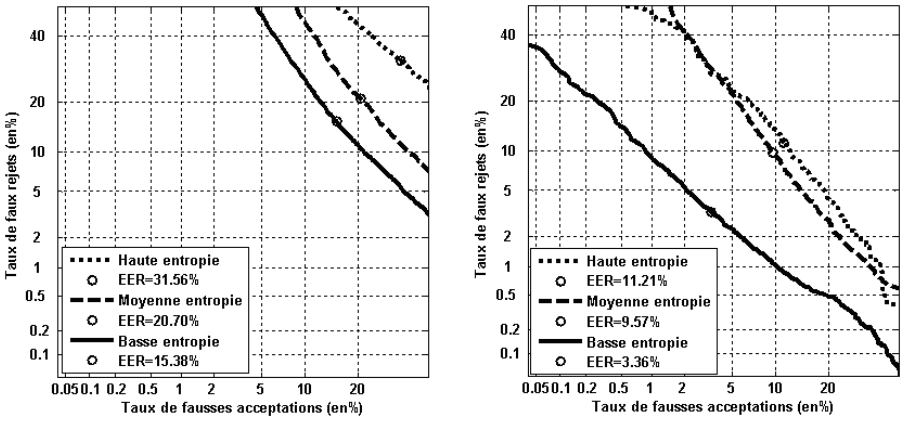

FIGURE 4.12: Courbes de performance avec les bonnes imitations (à gauche) et les imitations aléatoires (à droite) pour chaque catégorie de personnes de MCYT-100 avec le classifieur MMC.

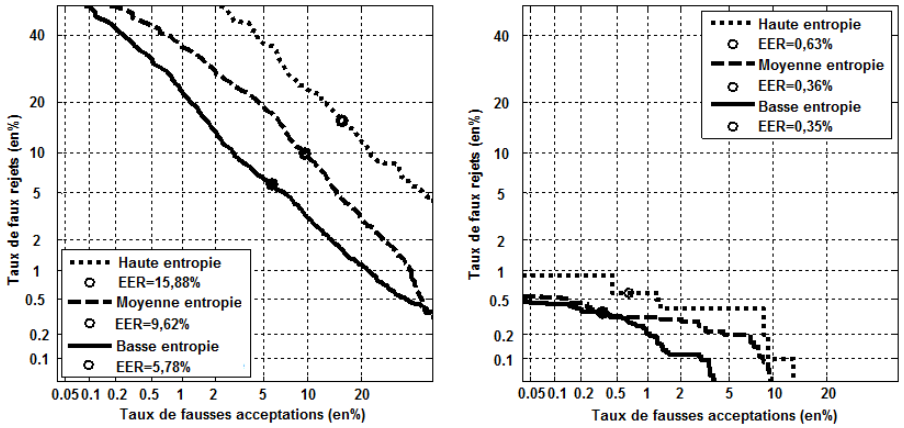

FIGURE 4.13: Courbes de performance avec les bonnes imitations (à gauche) et les imitations aléatoires (à droite) pour chaque catégorie de personnes de MCYT-100 avec le classifieur DTW.

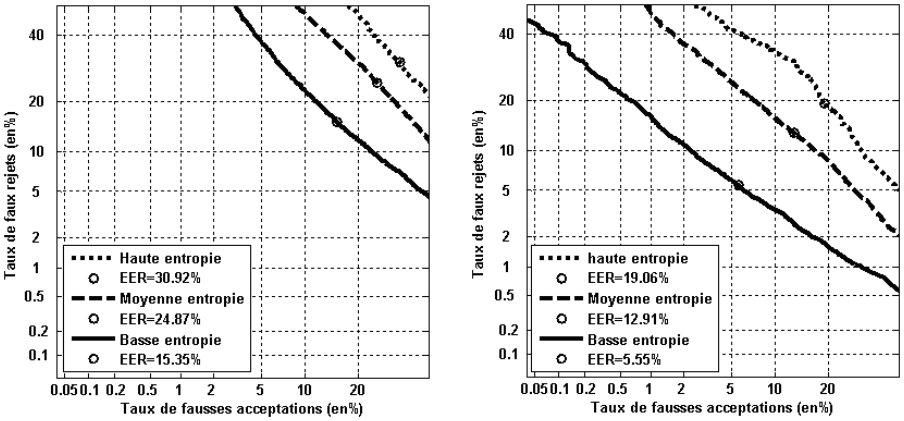

FIGURE 4.14: Courbes de performance avec les bonnes imitations (à gauche) et les imitations aléatoires (à droite) pour chaque catégorie de personnes de DS2-104 avec le classifieur MMC.

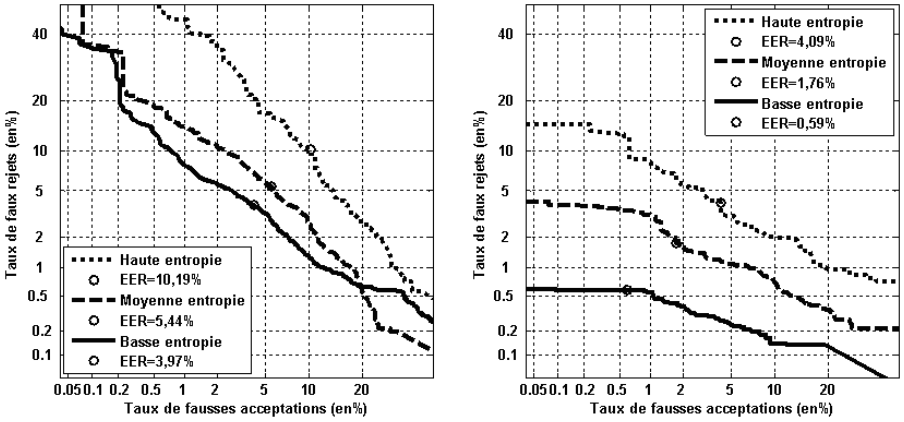

FIGURE 4.15: Courbes de performance avec les bonnes imitations(à gauche) et les imitations aléatoires (à droite) pour chaque catégorie de personnes de DS2-104 avec le classifieur DTW.

D'emblée, on constate que différents comportements en termes de performance émergent suivant la catégorie d'Entropie Personnelle considérée, et cela pour les deux classifieurs et les deux types d'imitations.

Il existe une différence significative dans les performances des classifieurs MMC et DTW entre les deux catégories extrêmes pour les deux types d'imitations : les meilleures performances sont obtenues avec la catégorie "Basse Entropie", celle contenant les signatures les plus longues, les plus complexes et les plus stables. A l'inverse, une dégradation importante des performances est observée sur les personnes appartenant à la catégorie "Haute Entropie", celle contenant les signatures les plus courtes, les plus simples et les plus instables. On remarque aussi qu'avec les deux classifieurs, la catégorie "Moyenne Entropie" donne des performances intermédiaires entre celles obtenues avec les deux catégories extrêmes, comme on pouvait s'y attendre.

Les Tableaux 4.4, 4.5, 4.6 et 4.7 montrent les performances des deux classifieurs MMC et DTW au point de fonctionnement *EER* par catégorie de personnes sur MCYT-100 et DS2-104. Les intervalles de confiance à 95% sont indiqués aussi pour montrer la pertinence des résultats.

Catégories de personnes	MCYT-100 (MMC)			
	Bonnes imitations		Imitations aléatoires	
	EER (%)	IC(95%)	EER (%)	IC(95%)
Haute Entropie	31,56	$\pm 3,572$	11,21	$\pm 1,621$
Moyenne Entropie	20,70	$\pm 1,391$	9,57	$\pm 0,831$
Basse Entropie	15,38	$\pm 0,653$	3,36	$\pm 0,289$

TABLEAU 4.4: Taux d'erreur à l'*EER* et intervalles de confiance par catégorie de personnes de MCYT-100 avec le classifieur MMC.

Catégories de personnes	MCYT-100 (DTW)			
	Bonnes imitations		Imitations aléatoires	
	EER (%)	IC(95%)	EER (%)	IC(95%)
Haute Entropie	15,88	$\pm 2,117$	0,63	$\pm 0,320$
Moyenne Entropie	9,62	$\pm 0,665$	0,36	$\pm 0,108$
Basse Entropie	5,78	$\pm 0,468$	0,35	$\pm 0,063$

TABLEAU 4.5: Taux d'erreur à l'*EER* et intervalles de confiance par catégorie de personnes de MCYT-100 avec le classifieur DTW.

On observe que le taux d'erreur *EER* diminue d'un facteur 2 au minimum lorsqu'on passe de la catégorie "Haute Entropie" à la catégorie "Basse Entropie". De surcroît, aux autres points de fonctionnement, cet écart de performance entre les deux catégories extrêmes est généralement maintenu pour les deux classifieurs, comme le montrent les courbes de performance sur les Figures 4.12, 4.13, 4.14 et 4.15.

Catégories de personnes	DS2-104 (MMC)			
	Bonnes imitations		Imitations aléatoires	
	EER (%)	IC(95%)	EER (%)	IC(95%)
Haute Entropie	30,92	$\pm 1,534$	19,06	$\pm 1,315$
Moyenne Entropie	24,87	$\pm 0,762$	12,91	$\pm 0,839$
Basse Entropie	15,35	$\pm 0,607$	5,55	$\pm 0,236$

TABLEAU 4.6: Taux d'erreur à l'*EER* et intervalles de confiance par catégorie de personnes de DS2-104 avec le classifieur MMC.

Catégories de personnes	DS2-104 (DTW)			
	Bonnes imitations		Imitations aléatoires	
	EER (%)	IC(95%)	EER (%)	IC(95%)
Haute Entropie	10,19	$\pm 0,448$	4,092	$\pm 0,448$
Moyenne Entropie	5,44	$\pm 0,403$	1,764	$\pm 0,146$
Basse Entropie	3,97	$\pm 0,220$	0,592	$\pm 0,038$

TABLEAU 4.7: Taux d'erreur à l'*EER* et intervalles de confiance par catégorie de personnes sur DS2-104 avec le classifieur DTW.

On peut donc conclure que le comportement des 3 catégories d'Entropie Personnelle en termes de performances reste stable pour différents classifieurs : une importante dégradation des performances est observée avec la catégorie "Haute Entropie" ; les meilleures performances sont obtenues avec la catégorie "Basse Entropie" ; la catégorie de transition celle à "Moyenne Entropie" donne des performances intermédiaires entre ces catégories extrêmes.

Ce résultat nous permet alors d'inférer qu'il y a des personnes "problématiques" qui sont de loin plus difficiles à reconnaître que d'autres : celles de la catégorie "Haute Entropie". Alors que les personnes de la catégorie "Basse Entropie" semblent être les plus fiables en termes de performances.

En comparaison avec la classification de la "Ménagerie Biométrique" présentée dans la Section 3.3 du Chapitre 3, on peut déduire que la catégorie "Haute Entropie" correspond à la catégorie des "*Goats*" contenant des personnes difficiles à reconnaître ; les catégories "Moyenne" et "Basse Entropie" correspondent à la catégorie des "*Sheeps*" contenant des personnes faciles à reconnaître. L'avantage indéniable de notre catégorisation est qu'elle est indépendante de toute étape de classification, à l'inverse de la "Ménagerie Biométrique" qui obtient les catégories pour un classifieur donné. Cependant, à ce niveau, nous ne trouvons que 2 catégories de personnes (dans la "Ménagerie Biométrique" il y a en tout 8 catégories) car la mesure d'Entropie Personnelle ne prend en compte que des données authentiques des personnes, alors que la "Ménagerie Biométrique" prend en compte des données authentiques ainsi que des imitations. De ce fait, les autres catégories de

la "Ménagerie Biométrique" n'apparaissent pas avec la mesure d'Entropie Personnelle. Nous verrons dans le Chapitre 5 comment trouver ces catégories à l'aide de la mesure d'Entropie Relative.

Pour un meilleur aperçu sur l'impact des catégories "Haute" et "Moyenne Entropie" sur les performances des systèmes de vérification, nous avons ordonné, dans le sens décroissant, les personnes appartenant à ces deux catégories en fonction de la valeur de leur Entropie Personnelle. Ensuite, en enlevant les premiers $x\%$ de ces personnes, nous avons calculé l'amélioration relative $\Delta(x)$ du taux d'erreur EER par rapport au taux d'erreur global (noté $\overline{EER}$) calculé sur toute la base de signatures, toutes catégories confondues :

$$\Delta(x) = \frac{\overline{EER} - EER(x)}{\overline{EER}} \tag{4.16}$$

où $EER\ (x)$ représente l'EER en considérant toute la base de signatures après avoir enlevé $x\%$ des personnes des catégories "Haute" et "Moyenne Entropie".

La Figure 4.16 illustre l'amélioration relative $\Delta(x)$ à l'EER en fonction du pourcentage de personnes enlevées de la catégorie "Haute" et "Moyenne Entropie". Et cela, pour les deux bases MCYT-100 et DS2-104 en utilisant les deux classifieurs DTW et MMC et en ne tenant compte que des bonnes imitations.

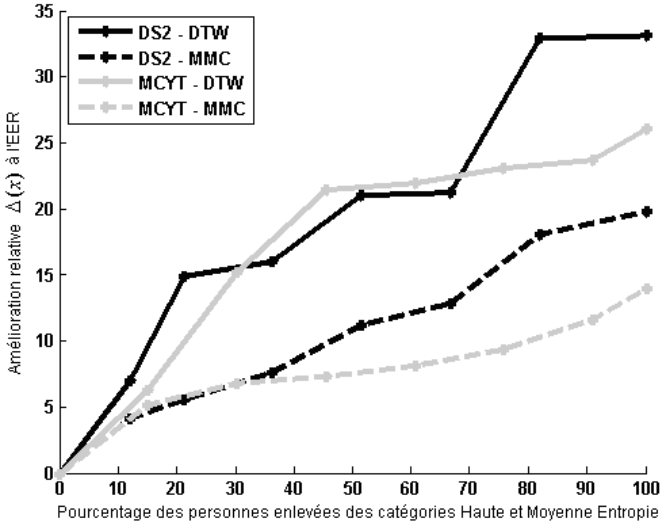

FIGURE 4.16: Amélioration relative $\Delta(x)$ à l'EER des classifieurs MMC et DTW sur MCYT-100 et DS2-104 avec les bonnes imitations, en supprimant $x\%$ de personnes contenues dans les catégories à "Haute" et "Moyenne Entropie".

On remarque que lors de la suppression progressive de personnes appartenant aux catégories "Haute" et "Moyenne Entropie", l'amélioration relative à l'EER $\Delta(x)$ augmente pour les deux classifieurs.

De plus, on observe que la principale amélioration est obtenue après la suppression des premières 15 personnes des catégories "Haute" et "Moyenne Entropie" pour MCYT-100 et des premières 17 personnes pour DS2-104. C'est à dire après la suppression de toutes les personnes de la catégorie "Haute Entropie" pour les deux bases, et de 10 personnes sur les 28 de la catégorie "Moyenne Entropie" pour MCYT-100 et de 10 personnes sur les 26 de la catégorie "Moyenne Entropie" pour DS2-104.

Base	Classifieur	Type d'imitations	$\overline{EER}(\%)$	$\Delta(100\%)$
MCYT-100	MMC	Bonnes imitations	17,87	13,94%
		imitations aléatoires	5,89	43,03%
	DTW	Bonnes imitations	7,81	26%
		imitations aléatoires	0,43	19,19%
DS2-104	MMC	Bonnes imitations	19,14	19,82%
		imitations aléatoires	8,58	35,38%
	DTW	Bonnes imitations	5,94	33,11%
		imitations aléatoires	1,21	51,12%

TABLEAU 4.8: Amélioration relative à l'*EER* des classifieurs MMC et DTW sur MCYT-100 et DS2-104, en supprimant toutes les personnes des catégories "Haute" et "Moyenne Entropie".

D'après le Tableau 4.8, on constate qu'une fois toutes les personnes appartenant aux catégories "Haute" et "Moyenne Entropie" retirées ($x = 100\%$), $\Delta(x)$ peut atteindre 26% pour les bonnes imitations et 43% pour les imitations aléatoires dans le cas de MCYT-100. Quant à la base DS2-104, $\Delta(x)$ peut atteindre 33% pour les bonnes imitations et 51% pour les imitations aléatoires. Une amélioration relative qui est loin d'être négligeable.

4.7 Conclusion

Dans ce chapitre, nous avons introduit notre mesure de qualité basée sur l'Entropie Personnelle appliquée aux signatures authentiques seulement. L'Entropie Personnelle a été calculée d'une part globalement sur toute la signature par le biais d'un Modèle à Mélange de Gaussiennes, et d'autre part localement sur des portions de la signature par le biais d'un Modèle de Markov Caché. Les deux modèles ont été tous les deux entraînés sur un ensemble de signatures authentiques de la personne. Nous avons trouvé que l'approche locale est mieux adaptée au contexte des signatures en-ligne pour obtenir une mesure d'Entropie garantissant une bonne catégorisation des données.

Nous avons ainsi obtenu des catégories de personnes cohérentes en termes d'aspect visuel des signatures, de complexité et de variabilité : la catégorie "Haute Entropie" contient des signatures courtes, simples et instables, ayant souvent l'aspect d'un paraphe. A l'opposé, la catégorie "Basse Entropie" contient des signatures longues, les plus complexes et les plus stables, se rapprochant parfois de l'écriture cursive. Entre ces deux catégories, la catégorie "Moyenne Entropie" correspond à une catégorie de transition entre les deux extrêmes, et contient des signatures plus longues que celles de la catégorie "Haute Entropie" et ont souvent l'aspect de paraphes complexes.

Puis, pour légitimer l'originalité de notre nouvelle mesure de qualité, nous l'avons comparée aux critères de qualité de l'état de l'art : nous avons montré que notre mesure d'Entropie Personnelle englobe à elle seule les 3 critères utilisés dans la littérature, à savoir, la complexité, la variabilité et la lisibilité.

Nous avons ensuite évalué les performances de deux classifieurs sur les personnes regroupées en 3 catégories dépendantes de leur Entropie Personnelle. Nous avons observé que différents comportements émergent suivant la catégorie de personne considérée. En effet, nous avons constaté qu'il y a des personnes de loin plus difficiles à caractériser et à reconnaître que d'autres : celles de la catégorie "Haute Entropie". Alors que les personnes de la catégorie "Basse Entropie" sont celles qui donnent les meilleures performances et semblent alors être les plus fiables en termes de sécurité.

Chapitre 5

Mesure de vulnérabilité des signatures authentiques aux imitations

Nous avons proposé dans le Chapitre 4 une mesure de qualité des signatures en-ligne, dénommée "Entropie Personnelle", opérant uniquement sur les signatures authentiques. Bien que cette mesure de qualité ait montré des résultats probants, elle reste néanmoins insuffisante pour prédire les performances des systèmes de vérification, mesurées en termes des deux types d'erreurs : les faux rejets et les fausses acceptations. En effet, pour avoir une mesure de qualité qui prenne en compte ces deux types d'erreurs, il est primordial que cette mesure de qualité tienne compte non seulement des signatures authentiques mais également des imitations.

Dans cette optique, nous proposons dans ce chapitre une nouvelle mesure de qualité des signatures en-ligne, basée sur le concept d'entropie relative, qui traite de la qualité des signatures authentiques confrontées aux imitations.

Après un rappel de la définition de l'entropie relative, appelée aussi divergence de Kullback-Leibler, nous verrons successivement comment l'entropie relative est calculée pour quantifier la vulnérabilité d'une signature aux imitations, pour ainsi aboutir à une mesure de qualité par personne dénommée "Entropie Relative Personnelle". Nous montrerons ensuite le lien existant entre cette nouvelle mesure de qualité basée sur l'entropie relative et notre précédente mesure basée sur l'entropie. Par la suite, nous confronterons la mesure d'Entropie Relative à notre précédente mesure d'Entropie et ce, en termes de catégories de personnes. Enfin, nous évaluerons l'impact de l'Entropie Relative sur les performances du système

de vérification basé sur la distance élastique, par rapport à la mesure d'Entropie Personnelle, sur différentes bases et pour différents types d'imitations.

5.1 Définition de l'entropie relative

Introduite en 1951, la divergence de Kullback-Leibler (appelée aussi entropie relative) doit son nom à *Solomon Kullback et Richard Leibler* [75], deux cryptanalystes américains. Cette divergence mesure la dissimilarité entre deux lois de probabilités p et q.

On considère une variable aléatoire Z de distribution de probabilité $p(z)$. Imaginons que, pour une raison ou une autre, on ait accès à une autre distribution $q(z)$, définie sur le même alphabet, pour décrire cette variable aléatoire Z. La divergence de Kullback-Leibler $D_{KL}(p||q)$ est alors définie par :

$$D_{KL}(p||q) = \sum_z p(z)\ log_2\ \frac{p(z)}{q(z)} \tag{5.1}$$

Cette quantité, exprimée en bits, mesure la divergence de $q(z)$ par rapport à $p(z)$. Elle permet de quantifier en nombre de bits l'écart entre ces deux distributions $q(z)$ et $p(z)$. Pour la notation, nous avons adopté celle qu'utilisent *Thomas et Cover* [18], où les arguments sont séparés par le signe "||".

La divergence de Kullback-Leibler s'interprète souvent comme un critère de dissimilarité entre une distribution et son estimation. Plus la valeur de la divergence est faible, plus grande est l'adéquation entre la distribution considérée et son estimation [18].

Cette divergence est toujours positive, elle est nulle si et seulement si les deux distributions sont identiques. Par ailleurs, elle n'est pas symétrique puisque $D_{KL}(p||q) \neq D_{KL}(q||p)$ et ne respecte pas l'inégalité triangulaire [18, 102]. La positivité de la divergence de Kullback-Leibler et le fait que pour $p = q$ on obtient la valeur "0" sont pour l'instant les deux seules et minces raisons qui justifient son emploi comme une mesure de proximité entre deux distributions. Bien qu'elle ne remplisse pas tous les axiomes d'une mesure de distance, il n'empêche que la divergence de Kullback-leibler est souvent appelée, par abus de langage, distance de Kullback-Leibler, mais elle n'est certainement pas une mesure de "distance" au sens mathématique du terme [18, 102].

Dans la littérature, la divergence de Kullback-Leibler est aussi appelée *entropie relative*, *information relative*, *information de discrimination*... En outre, on rencontre couramment la version symétrique de la distance de Kullback-Leibler [102] entre deux distributions $p(z)$ et $q(z)$, définie par :

$$D_{KL}(p,q) = \frac{1}{2}\left[D_{KL}(p||q) + D_{KL}(q||p)\right] \tag{5.2}$$

On utilisera cette version symétrisée de la divergence de Kullback-Leibler dans la suite de ce chapitre. Aussi, nous adopterons deux nominations : "Entropie Relative" pour être en phase avec l'ancienne mesure d'Entropie, et "divergence de Kullback-Leibler" afin de faciliter dans certains cas la compréhension en réfléchissant en termes de proximité.

5.2 Mesure d'Entropie Relative avec deux MMCs

De façon analogue à notre mesure d'Entropie Personnelle, et pour les mêmes raisons que celles évoquées précédemment dans la Section 4.3, dans ce chapitre aussi, les signatures sont représentées uniquement par les coordonnées *(x,y)*. Ceci nous permettra de confronter et de comparer nos deux mesures de qualité, à savoir : la mesure d'Entropie Relative Personnelle introduite dans ce chapitre et notre précédente mesure d'Entropie Personnelle.

L'Entropie Relative permet de mesurer l'écart entre deux distributions de probabilité définies sur un même Alphabet [18]. Une bonne estimation des distributions de probabilité est alors nécessaire pour un calcul précis de l'Entropie Relative.

Nous avons montré dans le Chapitre 4, qu'une estimation locale des densités de probabilité en se servant d'un Modèle de Markov Caché (MMC) [96] est une approche naturelle qui est très bien adaptée au signal de la signature. En effet, nous avons montré que lorsque l'on considère la signature comme un signal stationnaire par morceaux, correspondant aux réalisations de différentes variables aléatoires Z_i, on pouvait estimer avec précision par un MMC, les densités de probabilité de ces variables aléatoires, où les réalisations sont des couples de coordonnées *(x,y)* sur une portion donnée S_i de la signature [43]. Ces portions sont générées par l'algorithme de Viterbi et correspondent aux états du MMC [96].

Dans le Chapitre 4, pour le calcul de l'Entropie Personnelle, l'estimation de densité locale est réalisée sur un ensemble de signatures authentiques, en entraînant un MMC sur ces signatures avec un nombre d'états personnalisé. Nous avons précisé que pour avoir une bonne estimation de densités, il est nécessaire d'avoir au moins 10 signatures authentiques par personne [43].

Dans le même esprit, pour le calcul de l'Entropie Relative, nous estimons pour chaque personne les densités de probabilités locales sur un ensemble de 10 signatures authentiques de la personne, mais aussi sur un ensemble de 10 bonnes imitations associées à cette personne. Et cela, par le biais de deux MMCs différents, l'un appris sur les 10 signatures authentiques et l'autre sur les 10 bonnes imitations associées à cette même personne. Le nombre d'états des deux MMCs n'est certainement pas le même car il dépend de la longueur totale des signatures d'apprentissage (d'après l'équation 4.9).

Par conséquent, pour chaque personne, nous construisons deux Modèles de Markov Cachés ; dont l'un est entraîné sur 10 signatures authentiques et l'autre sur 10 bonnes imitations. Puis, nous estimons localement les densités de probabilités p et q sur l'ensemble de ces signatures authentiques et de ces bonnes imitations respectivement. Ensuite, pour chaque personne, nous calculons la version symétrisée de l'Entropie Relative entre les deux distributions sur un même ensemble d'alphabet. Cet alphabet représente l'ensemble des couples *(x,y)* appartenant à chaque signature, parmi 10 signatures authentiques de la personne et 10 bonnes imitations qui lui sont associées, différentes de celles utilisées pour estimer les densités de probabilités avec les deux MMCs.

Ainsi, sur chaque signature authentique sig_i parmi les 10, et sur chaque bonne imitation im_j parmi les 10, nous calculons respectivement suivant la version symétrisée :

$$D_{sig_i}(p,q) = \frac{1}{2}\left[\sum_{z\in sig_i} p(z)\ log_2\ \frac{p(z)}{q(z)} + \sum_{z\in sig_i} q(z)\ log_2\ \frac{q(z)}{p(z)}\right] \tag{5.3}$$

$$D_{im_j}(p,q) = \frac{1}{2}\left[\sum_{z\in im_j} p(z)\ log_2\ \frac{p(z)}{q(z)} + \sum_{z\in im_j} q(z)\ log_2\ \frac{q(z)}{p(z)}\right] \tag{5.4}$$

où $p(z)$ est la valeur de la densité de probabilité locale au point $z = (x, y)$ appartenant à la portion courante suivant le MMC appris sur les signatures authentiques, et $q(z)$ est la valeur de la densité de probabilité locale au point $z = (x, y)$ appartenant à la portion courante suivant le MMC appris sur les bonnes imitations. Les

deux sommes sont effectuées sur tous les points de la $i^{\text{ème}}$ signature authentique (sig_i) et de la $j^{\text{ème}}$ bonne imitation (im_j). Ce processus de calcul est illustré sur les Figures 5.1 et 5.2.

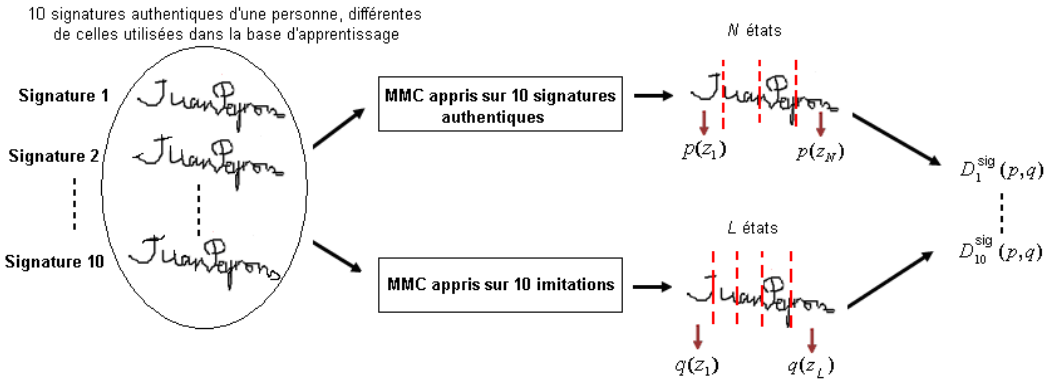

FIGURE 5.1: Calcul de la divergence de Kullback-Leibler pour une personne donnée entre ses signatures authentiques et ses imitations, sur ses 10 autres signatures authentiques, après apprentissage des deux MMCs.

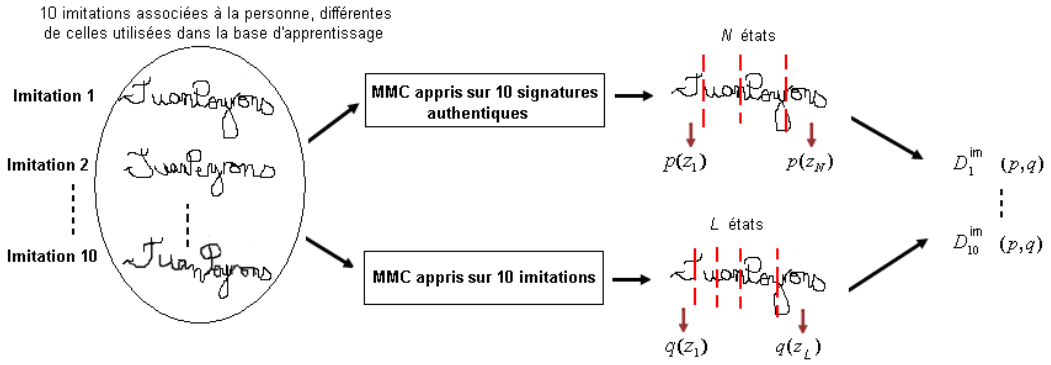

FIGURE 5.2: Calcul de la divergence de Kullback-Leibler pour une personne donnée entre ses signatures authentiques et ses imitations, sur ses 10 autres imitations, après apprentissage des deux MMCs.

A noter que, comme les densités de probabilité $p(z)$ et $q(z)$ sont estimées localement, elles vont effectivement changer suivant la portion visitée par chaque point de la signature et cela pour chaque MMC.

Finalement, afin de calculer l'Entropie Relative Personnelle $D_p(p,q)$, nous moyennons tout d'abord séparément $D_{sig_i}(p,q)$ sur les 10 signatures authentiques de la personne considérée et $D_{im_j}(p,q)$ sur les 10 bonnes imitations associées à cette personne :

$$D_{sig}(p,q) = \frac{1}{10}\sum_{i=1}^{10} D_{sig_i}(p,q) \tag{5.5}$$

$$D_{im}(p,q) = \frac{1}{10}\sum_{j=1}^{10} D_{im_j}(p,q) \tag{5.6}$$

Puis, nous moyennons ces deux valeurs avec une simple moyenne arithmétique :

$$D_p(p,q) = \frac{1}{2}\left[D_{sig}(p,q) + D_{im}(p,q)\right] \tag{5.7}$$

L'Entropie Relative Personnelle $D_p(p,q)$ mesure ainsi pour une personne donnée la divergence de Kullback-Leibler entre les lois de probabilités locales de ses signatures authentiques et des imitations qui lui sont associées. De la sorte, plus la divergence est faible, plus les imitations sont réputées proches des signatures authentiques, et plus la personne est considérée vulnérable aux attaques (c.à.d. sa signature est facile à imiter) ; plus la divergence est grande, plus la personne est considérée robuste aux attaques (sa signature est difficile à imiter).

Toutefois, il est vrai aussi que si la divergence de Kullback-Leibler entre les lois de probabilités locales des signatures authentiques et de leurs imitations est faible, cela pourrait dépendre lorsque les signatures imitées sont réalisées par le même imposteur, de son don en tant qu'imitateur. Cependant, comme nous le verrons dans la partie expérimentale de ce chapitre, nous avons utilisé pour chaque personne des imitations provenant de différents imposteurs pour que la valeur de l'Entropie Relative calculée soit indépendante d'un imposteur bien spécifique.

5.3 Entropie Relative comparée à l'Entropie

Dans le contexte de la théorie de l'information, la divergence de Kullback-Leibler est aussi appelée écart entropique ou entropie relative [18] à cause de son lien avec l'entropie. En fait, l'expression de l'entropie est intégrée dans celle de l'entropie relative et correspond à son premier membre, comme montré ci-dessous :

$$\begin{aligned} D_{KL}(Z) = D(p\|q) &= \sum_z p(z)\ log_2\ \frac{p(z)}{q(z)} \\ &= \sum_z p(z)\ log_2\ p(z)\ -\ \sum_z p(z)\ log_2\ q(z) \\ &= H(Z)\ -\ \sum_z p(z)\ log_2\ q(z) \end{aligned} \tag{5.8}$$

Par ailleurs, dans le contexte de notre travail de thèse, l'Entropie Personnelle, par construction, est une mesure dépendante de la personne qui caractérise les densités de probabilités locales des signatures authentiques de la personne. En revanche, l'Entropie Relative Personnelle, comme nous l'avons calculée, est une

mesure *discriminante* dépendante de la personne, qui indique la divergence entre les densités de probabilité locales des signatures authentiques de la personne et celles des imitations associées à cette personne.

En conséquence, l'Entropie Relative paraît naturellement mieux adaptée aux systèmes biométriques confrontés constamment au problème de discrimination entre les données authentiques et les données imitations. C'est ce que nous allons démontrer dans la partie suivante.

5.4 Entropie Relative et Entropie en termes de catégories de personnes

Dans cette partie, nous allons tout d'abord analyser le comportement des catégories, générées par la mesure d'Entropie Personnelle présentée dans le Chapitre 4, en termes de l'autre mesure de qualité, à savoir l'Entropie Relative Personnelle. Ceci nous permettra en premier temps de voir s'il existe un lien entre la mesure d'Entropie Personnelle et la mesure d'Entropie Relative Personnelle en termes de catégories de personnes. Nous nous intéresserons ultérieurement à la catégorisation de personnes suivant leur valeur d'Entropie Relative Personnelle, et nous comparerons ensuite ces catégories à celles obtenues avec l'Entropie Personnelle. Le but étant d'analyser l'impact de notre nouvelle mesure d'entropie discriminante sur la catégorisation de personnes.

Les expériences dans ce chapitre sont effectuées sur les deux bases de signatures en-ligne MCYT-100 et Philips. La base MCYT-100 a été choisie car elle a déjà été utilisée pour le calcul de l'Entropie Personnelle dans le Chapitre 4, ce qui nous permettra de comparer les deux mesures de qualité. De plus, pour un meilleur aperçu visuel des résultats, on peut montrer des exemples de signatures de MCYT-100, déjà publiés dans [3, 32, 91]. Quant à la base Philips, elle a été choisie car elle contient des imitations de différents types. Ceci nous permettra d'étudier la vulnérabilité des personnes lorsque leur signature est confrontée à des imitations statiques et dynamiques. Les imitations professionnelles de la base Philips ne sont pas utilisées dans ce chapitre, car elles ne sont pas disponibles pour toutes les personnes de la base.

5.4.1 Catégories de personnes d'Entropie en termes d'Entropie Relative

5.4.1.1 Résultats sur la base MCYT-100

Nous avons représenté sur la Figure 5.3 notre nouvelle mesure d'Entropie Relative Personnelle en fonction de l'Entropie Personnelle, sur les personnes de MCYT-100 déjà classées en 3 catégories suivant leur valeur d'Entropie Personnelle. Sur cette figure, nous avons reporté des exemples de signatures représentant chaque catégorie de personnes afin de donner un aperçu visuel sur leur contenu. A noter que ces exemples de signatures ont déjà été publiés dans [3, 91, 32] et ont fait l'objet d'une autorisation de la part de leurs auteurs.

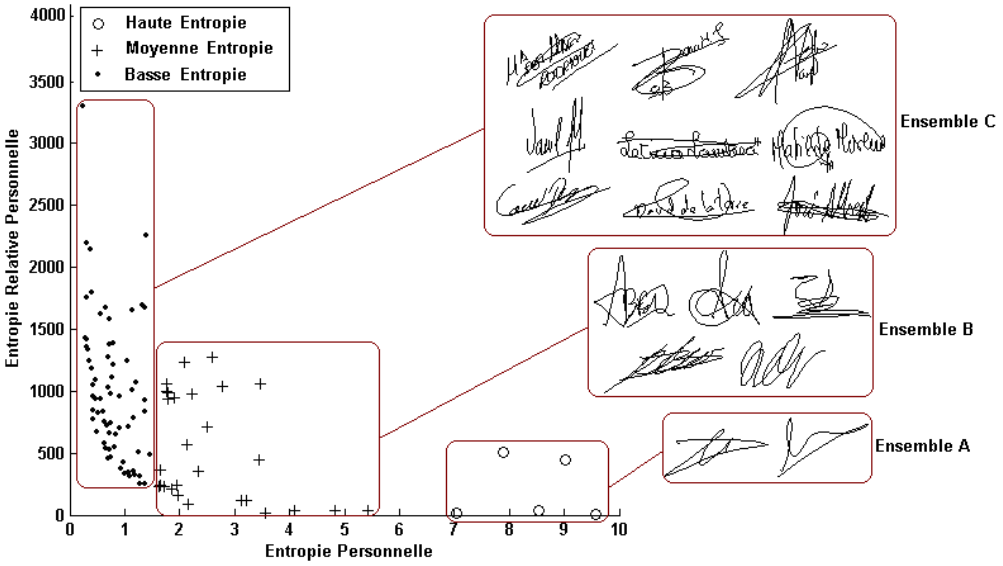

FIGURE 5.3: Entropie Relative Personnelle en fonction de l'Entropie Personnelle pour chaque catégorie de personnes de MCYT-100 : (o) Haute, (+) Moyenne et (.) Basse Entropie Personnelle.

D'après la Figure 5.3, on remarque l'existence d'une relation inverse entre l'Entropie et l'Entropie Relative, confirmée par la matrice de corrélation donnée dans le Tableau 5.1, mesurant la corrélation entre l'Entropie et l'Entropie Relative pour toutes les personnes de la base MCYT-100.

De fait, on observe que les imitations sont très proches des signatures authentiques, en termes de divergence de Kullback-Leibler, pour les personnes appartenant à la catégorie "Haute Entropie", dont les signatures sont les plus variables et les moins complexes (voir l'ensemble *A* Figure 5.3). Il apparaît donc clairement que

ces personnes sont les plus vulnérables aux attaques. Dans le cas de MCYT-100, les attaques sont de bonnes imitations statiques [91].

Variables	Entropie Personnelle	Entropie Relative Personnelle
Entropie Personnelle	1	-0,45
Entropie Relative Personnelle	-0,45	1

TABLEAU 5.1: Matrice de corrélation entre l'Entropie et l'Entropie Relative sur la base MCYT-100.

D'autre part, les personnes appartenant à la catégorie "Basse Entropie", dont les signatures sont les plus longues, les plus stables et les plus complexes (voir l'ensemble C Figure 5.3), présentent des grandes valeurs de divergence de Kullback-Leibler entre leurs signatures authentiques et leurs imitations. Il apparaît donc clairement que ces personnes sont les moins vulnérables aux attaques (bonnes imitations statiques).

En ce qui concerne les personnes de la catégorie "Moyenne Entropie" (voir l'ensemble B Figure 5.3), on constate qu'elles présentent des valeurs de divergence de Kullback-Leibler intermédiaires entre celles des deux catégories extrêmes.

Ces premiers résultats rejoignent nos constatations évoquées à la fin du Chapitre 4 lors de l'évaluation des performances des systèmes de vérification, à savoir : les personnes de la catégorie "Haute Entropie" sont considérées comme des signataires "problématiques", difficiles à caractériser et à reconnaître, alors que les personnes de la catégorie "Basse Entropie" sont considérées comme les plus fiables en termes de performances.

5.4.1.2 Résultats sur la base Philips

La Figure 5.4 illustre l'Entropie Relative en fonction de l'Entropie sur les catégories de personnes de la base Philips générées avec l'Entropie Personnelle. Comme la base Philips contient différents types d'imitations, nous avons alors considéré deux cas de figure : dans le premier cas, l'Entropie Relative est calculée en considérant uniquement les imitations "améliorées chez soi, HI" (Figure 5.4.a) ; dans le deuxième cas, elle est calculée en considérant uniquement les imitations "par-dessus l'épaule, SH" (Figure 5.4.b). Pour rappel, les imitations "HI" sont des imitations statiques, alors que les imitations "SH" sont des imitations dynamiques (se reporter à la description de la base Philips dans la Section 2.4.3). Pour des raisons de confidentialité, nous ne pouvons pas montrer des exemples de signatures de la base Philips.

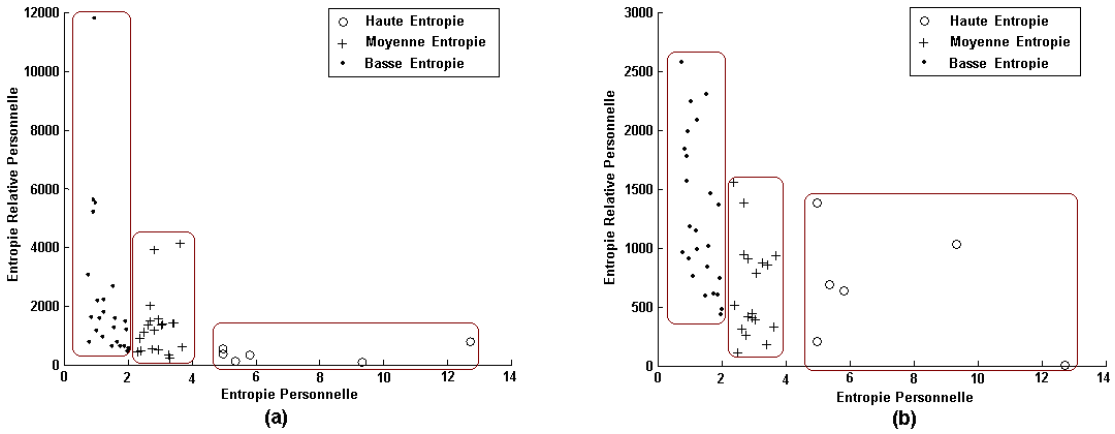

FIGURE 5.4: Entropie Relative Personnelle, calculée en considérant les imitations (a) *HI* et (b) *SH*, en fonction de l'Entropie Personnelle pour chaque catégorie de personnes de la base Philips : (o) Haute, (+) Moyenne et (.) Basse Entropie Personnelle.

Comme pour MCYT-100, on observe sur la Figure 5.4 une relation inverse entre l'Entropie et l'Entropie Relative sur la base Philips pour les deux types d'imitations.

D'autre part, lorsque l'on compare les valeurs de l'Entropie Relative obtenues avec les imitations statiques (Figure 5.4.a) à celles obtenues avec les imitations dynamiques (Figure 5.4.b), on constate que les valeurs de la divergence de Kullback-Leibler sont 4 fois plus élevées avec les imitations "HI" : un maximum de 12000 est atteint avec les imitations statiques (HI) contre 3000 avec les imitations dynamiques (SH). Cela revient à dire que les imitations "SH" sont 4 fois plus proches des signatures authentiques que les imitations "HI".

Ce résultat est très intéressant et confirme bien que les imitations "SH", qui sont des imitations de la dynamique de la signature d'origine, sont meilleures que les imitations "HI", qui sont des imitations de l'image de la signature d'origine. Par conséquent, la différence de qualité des imitations est donc bien quantifiée par notre mesure d'Entropie Relative Personnelle.

Jusqu'à présent, notre nouvelle mesure de qualité des signatures authentiques qui prend en compte leur vulnérabilité aux attaques conduit à des résultats cohérents sur deux bases différentes et pour différents types d'imitations. En se basant sur ces premiers résultats, cette nouvelle mesure de qualité basée sur l'Entropie Relative soulève inévitablement les questions suivantes : peut-on catégoriser les personnes de la base suivant leur valeur d'Entropie Relative ? Si oui, ces catégories

ont-elles un aspect similaire à celles générées avec l'Entropie Personnelle ? Sont-elles homogènes comme celles générées avec l'Entropie Personnelle ? Répondre à ces questions est notre objectif dans la section suivante.

5.4.2 Catégories de personnes d'Entropie Relative

Afin d'étudier l'impact de notre nouvelle mesure d'Entropie Relative sur la catégorisation de personnes, nous avons appliqué, tout comme pour l'Entropie Personnelle, une classification hiérarchique ascendante [89] sur les valeurs d'Entropie Relative Personnelle des bases MCYT-100 et Philips. Puis, nous avons procédé à l'étude du nombre de catégories optimal en faisant appel aux différents indices de validité [45], comme expliqué dans l'Annexe A. Nous avons trouvé que 3 catégories est le choix optimal du nombre de catégories. Ainsi, 3 nouvelles catégories ont émergé :

- catégorie "Haute Entropie Relative", contenant des signatures "peu vulnérables" aux attaques,

- catégorie "Moyenne Entropie Relative", contenant des signatures "moyennement vulnérables" aux attaques,

- catégorie "Basse Entropie Relative", contenant des signatures "très vulnérables" aux attaques.

5.4.2.1 Résultats sur la base MCYT-100

La Figure 5.5 montre l'Entropie Relative Personnelle en fonction de l'Entropie Personnelle pour les personnes de la base MCYT-100, classées en 3 nouvelles catégories suivant leur valeur d'Entropie Relative Personnelle. Afin de mettre en évidence le contenu de chaque catégorie, nous avons aussi reporté sur cette figure des exemples de signatures représentant chaque catégorie. A noter que ces exemples de signatures sont les mêmes que ceux utilisés sur la Figure 5.3 afin de distinguer les changements qui peuvent se produire dans les catégories avec notre nouvelle mesure de qualité.

On observe tout d'abord que la répartition des personnes en 3 catégories suivant leur valeur d'Entropie Relative Personnelle est différente de celle observée

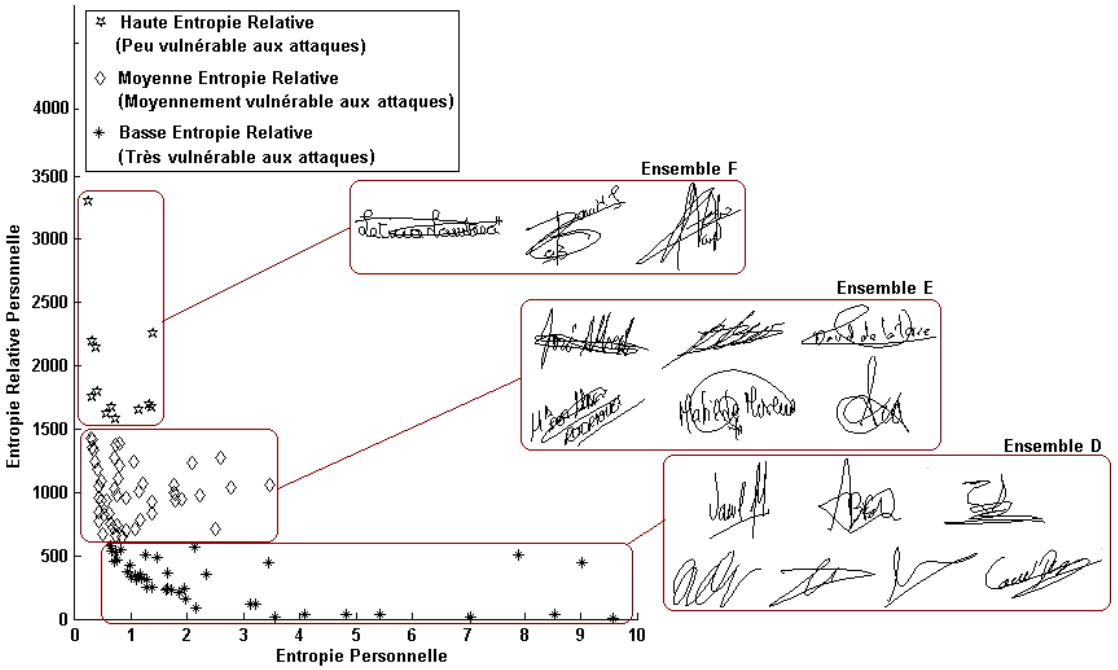

FIGURE 5.5: Entropie Relative Personnelle en fonction de l'Entropie Personnelle pour les nouvelle catégories de personnes de la base MCYT-100 : (☆) Haute, (◇) Moyenne et (∗) Basse Entropie Relative Personnelle.

sur la Figure 5.3, où les catégories sont générées avec la mesure d'Entropie Personnelle. Cette différence dans la répartition des personnes suivant les deux mesures de qualité est illustrée dans le Tableau 5.2.

Pourcentage de personnes par catégorie	Entropie Personnelle	Entropie Relative Personnelle
Haute	5% (Ensemble A)	40% (Ensemble D)
Moyenne	28% (Ensemble B)	48% (Ensemble E)
Basse	67% (Ensemble C)	12% (Ensemble F)

TABLEAU 5.2: Distribution de personnes de MCYT-100 dans chaque catégorie d'Entropie Personnelle et d'Entropie Relative Personnelle.

Certes, la catégorie "Basse Entropie Relative", contenant les signatures les plus vulnérables aux imitations (Ensemble *D* sur la Figure 5.5), renferme des signatures courtes, simples et variables, les mêmes que celles contenues dans la précédente catégorie "Haute Entropie" (Ensemble *A* sur la Figure 5.3). Ainsi, les personnes de la catégorie "Haute Entropie", correspondant aux signataires "problématiques", sont toutes jugées très vulnérables aux attaques avec la mesure d'Entropie Relative. Néanmoins, cette nouvelle catégorie "Basse Entropie Relative" contient aussi d'autres signatures plus longues et plus complexes. En effet, elle contient un plus grand pourcentage de personnes que la précédente catégorie

"Haute Entropie" : 40% contre 5% de la base MCYT-100 respectivement, comme montré dans le Tableau 5.2.

A contrario, la catégorie "Haute Entropie Relative", contenant les signatures les moins vulnérables aux imitations (Ensemble F sur la Figure 5.5), est en fait un sous-groupe de la précédente catégorie "Basse Entropie" (Ensemble C sur la Figure 5.3), et qui a la particularité d'être plus homogène que la catégorie "Basse Entropie". En effet, la nouvelle catégorie "Haute Entropie Relative" ne contient que des signatures cursives, les plus complexes, les plus stables, et en même temps les moins vulnérables aux attaques ; ces signatures ne représentent que 12% de toute la base MCYT-100 et 17,91% de la catégorie "Basse Entropie" (12 parmi 67, voir le Tableau 5.2, et en particulier la Figure 5.6).

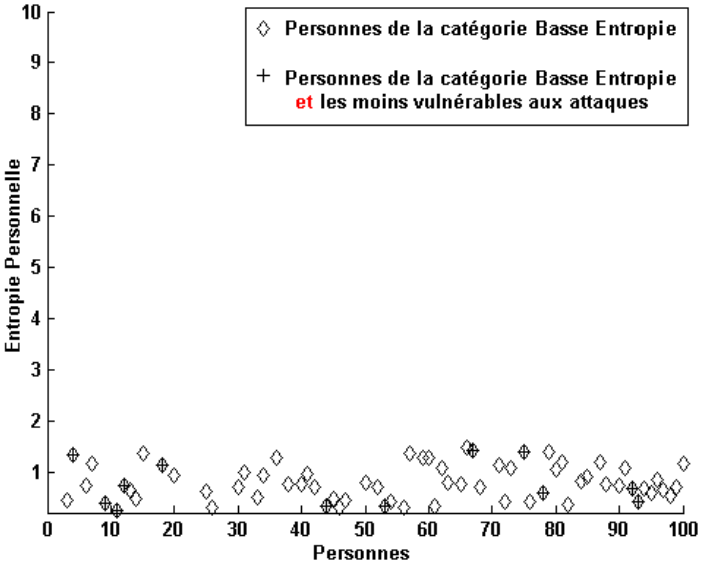

FIGURE 5.6: Valeurs d'Entropie Personnelle des personnes de MCYT-100 appartenant uniquement à la catégorie "Basse Entropie" représentées par ◇. Parmi elles, les personnes les plus robustes aux imitations, représentant 17,91% de la catégorie "Bassse Entropie" sont désignées par ⊕.

Dès lors, certaines personnes de la catégorie "Basse Entropie" (Ensemble C sur la Figure 5.3) se retrouvent maintenant dans les catégories "Moyenne Entropie Relative" (Ensemble E sur la Figure 5.5) ou "Basse Entropie Relative" (Ensemble D sur la Figure 5.3). En effet, 37 personnes parmi les 67 personnes de la catégorie "Basse Entropie" et 18 personnes parmi ces 67 se retrouvent maintenant dans la catégorie "Moyenne" et "Basse" Entropie Relative respectivement. C'est-à-dire qu'en considérant en plus le critère de vulnérabilité aux attaques, ces personnes sont jugées moyennement vulnérables voire très vulnérables aux attaques.

Ce résultat signifie que certaines signatures très complexes et stables ayant souvent l'aspect d'écriture cursive peuvent être "moyennement vulnérables" ou "très vulnérables" aux attaques. Ceci est expliqué, d'une part par le fait que certaines signatures de la catégorie "Basse Entropie", bien qu'elles soient assez complexes, restent tout de même faciles à imiter ; et d'autre part, par le fait que les imitations associées à une personne ne sont pas faites par un même et unique imposteur, et certains imposteurs sont tout simplement meilleurs imitateurs que d'autres. En effet, le protocole d'acquisition de MCYT stipule que la personne n est imitée par les personnes *(n+1)* à *(n+5)*. La Figure 5.7 montre des exemples de signatures qui ont changé de catégorie suivant leur valeur d'Entropie Relative.

FIGURE 5.7: Exemples de signatures de MCYT-100 (a) à "Moyenne Entropie" jugées très vulnérables aux imitations, (b) à "Basse Entropie" jugées très vulnérables aux imitations, et (c) à "Basse Entropie" jugées moyennement vulnérables aux imitations.

Par ailleurs, on remarque sur la Figure 5.6, que les signatures les moins vulnérables aux attaques, représentées par ⊕, ne sont pas nécessairement celles dont les valeurs d'Entropie Personnelle sont les plus faibles. Cette observation est très pertinente et est conforme avec nos attentes puisque d'après nos résultats, la mesure d'Entropie Relative Personnelle comprend non seulement les 3 critères usuels de la littérature comme notre mesure d'Entropie Personnelle (complexité, stabilité et lisibilité de la signature), mais aussi un quatrième critère qui est la vulnérabilité des signatures aux attaques.

5.4.2.2 Résultats sur la base Philips

La Figure 5.8 montre l'Entropie Relative Personnelle en fonction de l'Entropie Personnelle pour les personnes de la base Philips, classées en 3 nouvelles catégories suivant leur valeur d'Entropie Relative, en considérant séparément les deux types d'imitations : "améliorées chez soi, HI" et "par-dessus l'épaule, SH".

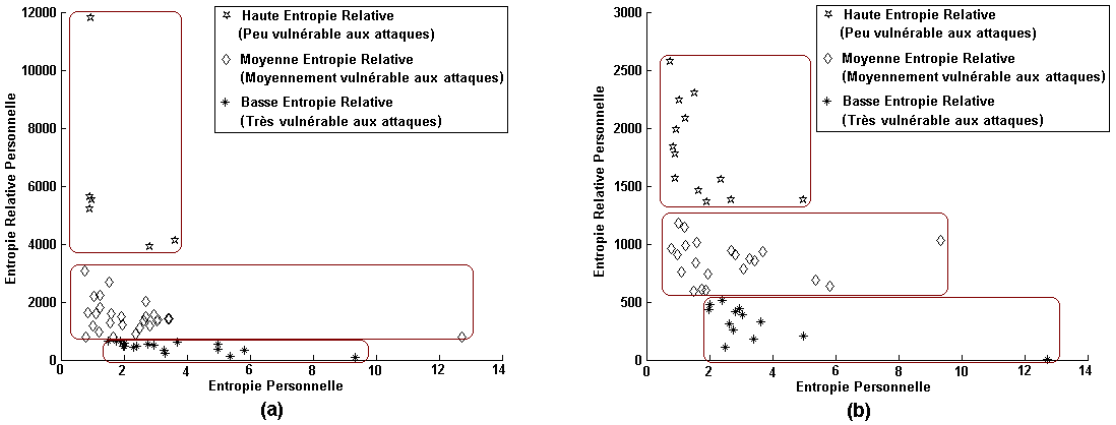

FIGURE 5.8: Entropie Relative Personnelle, calculée en considérant les imitations (a) *HI* et (b) *SH*, en fonction de l'Entropie Personnelle pour les nouvelles catégories de personnes de la base Philips : (☆) Haute, (◇) Moyenne et (∗) Basse Entropie Relative Personnelle.

Comme pour MCYT-100, on observe que la répartition des personnes de la base Philips en 3 catégories suivant leur valeur d'Entropie Relative Personnelle est différente de celle observée sur la Figure 5.4, où les catégories sont générées avec la mesure d'Entropie Personnelle, comme montré dans le Tableau 5.3.

Catégories de personnes	Entropie	Entropie Relative	
		imitations statiques	imitations dynamiques
Haute	11,76%	35,30%	30,62%
Moyenne	39,22%	52,94%	42,85%
Basse	49,02%	11,76%	26,53%

TABLEAU 5.3: Distribution des personnes de la base Philips dans chaque catégorie d'Entropie et d'Entropie Relative avec les deux types d'imitations (statiques et dynamiques).

Cependant, on constate sur la base Philips que certains comportements des catégories générées avec l'Entropie Relative sont différents de ceux observés dans la précédente Section 5.4.2.1 sur la base MCYT-100.

En effet, la catégorie "Haute Entropie Relative" n'est pas tout à fait un sous-ensemble de la catégorie "Basse Entropie" comme observé sur la base MCYT-100. En fait, en considérant les imitations statiques de la base Philips, la catégorie "Haute Entropie Relative", contenant des signatures complexes, stables, cursives et les moins vulnérables aux attaques, comporte 4 personnes de l'ancienne catégorie "Basse Entropie" et 2 personnes de l'ancienne catégorie "Moyenne Entropie". Ce résultat montre clairement que l'Entropie Relative s'emploie à trouver un bon compromis entre les 3 critères usuels de la littérature (complexité , variabilité et lisibilité) et la vulnérabilité des signatures aux attaques.

Aussi, une autre différence de comportement par rapport à MCYT-100 est à noter : parmi les 6 personnes problématiques de la catégorie “Haute Entropie”, il existe une personne qui est jugée moyennement vulnérable aux imitations statiques, et 3 personnes qui sont jugées moyennement vulnérables aux imitations dynamiques. Ce résultat ne fait que renforcer nos précédents résultats vis-à-vis de la bonne qualité des imitations dynamiques.

Par ailleurs, on remarque aussi que la distribution des personnes dans les 3 catégories d'Entropie Relative est nettement différente selon le type d'imitations (HI et SH). Cela induit un résultat très intéressant : les signatures authentiques qui sont vulnérables aux imitations statiques (HI) ne sont pas nécessairement vulnérables aux imitations dynamiques (SH), et vice-versa. En effet, seulement 7 personnes de la base Philips sont les plus vulnérables aux deux types d'imitations, et seulement 3 personnes de la base Philips sont les moins vulnérables aux deux types d'imitations.

5.4.3 Catégories de personnes générées par fusion d'Entropie et d'Entropie Relative

Pour se rendre mieux compte du lien existant entre l'Entropie et l'Entropie Relative en termes de catégories de personnes, nous avons effectué uniquement sur la base MCYT-100 une classification hiérarchique ascendante bidimensionnelle [89], à la fois sur les valeurs d'Entropie et les valeurs d'Entropie Relative simultanément.

La Figure 5.9 montre l'Entropie Relative en fonction de notre précédente mesure d'Entropie, pour les personnes de la base MCYT-100, en illustrant les catégories obtenues, après étude du nombre optimal de catégories [45]. Comme montré sur la Figure 5.9, on obtient 6 catégories de personnes suivant leurs valeurs d'Entropie Personnelle et d'Entropie Relative Personnelle, en se servant de la classification hiérarchique bidimensionnelle. Pour une meilleure illustration des catégories obtenues, nous avons reporté sur la Figure 5.9 des exemples de signatures représentant chaque catégorie. Ces exemples de signatures sont les mêmes que ceux utilisés dans la Section 5.4.1.1 (pour les catégories d'Entropie) et la Section 5.4.2.1 (pour les catégories d'Entropie Relative).

On observe que la catégorisation bidimensionnelle permet de sélectionner deux catégories extrêmes : la première catégorie (Ensemble A, Figure 5.9) correspond à la catégorie “Haute Entropie”, et contient les personnes “problématiques”

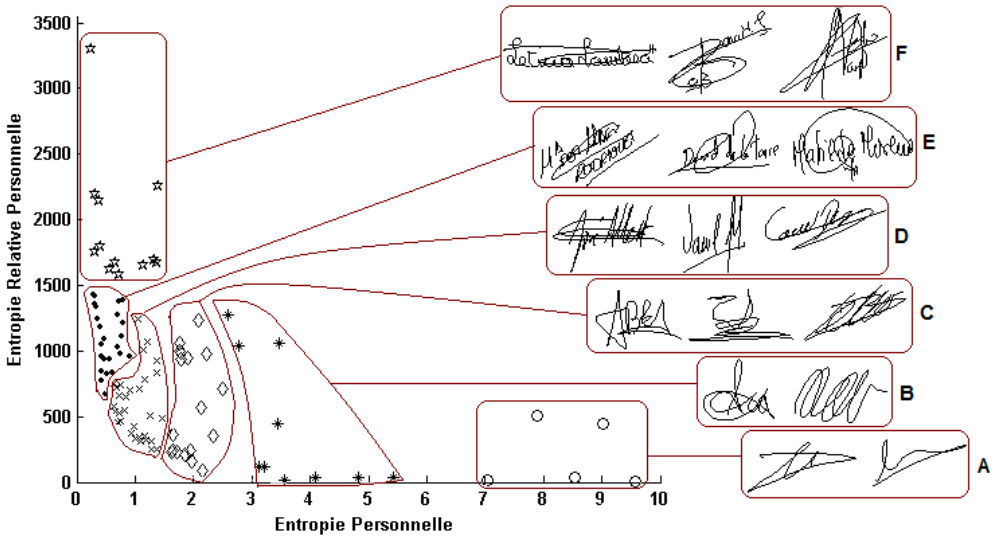

FIGURE 5.9: Entropie Relative Personnelle en fonction de l'Entropie Personnelle pour chaque catégorie de personnes de MCYT-100 générées par une classification hiérarchique 2D appliquées aux valeurs d'Entropie et d'Entropie Relative.

dont les signatures sont courtes, simples, très variables et très vulnérables aux attaques ; la deuxième catégorie extrême (Ensemble F, Figure 5.9) correspond à la catégorie "Haute Entropie Relative", et contient les signatures cursives, très complexes, très stables, et les moins vulnérables aux attaques. Entre ces deux catégories extrêmes A et F, on observe 4 catégories intermédiaires qui montrent une variation régulière des signatures en termes de complexité, de stabilité, de lisibilité et de vulnérabilité aux attaques.

Le Tableau 5.4 montre la distribution des personnes de MCYT-100 en chaque nouvelle catégorie générée par la classification bidimensionnelle.

Catégories	Pourcentage de personnes
$1^{ère}$ catégorie (≡ à la catégorie "Haute Entropie")	5%
$2^{ème}$ catégorie	10%
$3^{ème}$ catégorie	18%
$4^{ème}$ catégorie	31%
$5^{ème}$ catégorie	24%
$6^{ème}$ catégorie (≡ à la catégorie "Haute Entropie Relative")	12%

TABLEAU 5.4: Distribution des personnes de MCYT-100 dans chaque catégorie générée par fusion de l'Entropie et l'Entropie Relative.

Par rapport à la catégorisation de personnes basée soit sur l'Entropie Personnelle, soit sur l'Entropie Relative Personnelle, on constate qu'en combinant les deux mesures de qualité, la classification bidimensionnelle des personnes de MCYT-100 :

- n'affecte pas la catégorie "Haute Entropie" (Ensemble A, Figure 5.9), contenant les personnes "problématiques", dont les signatures sont les plus variables, les moins complexes et les plus vulnérables aux imitations.

- divise la catégorie "Moyenne Entropie" en 2 sous-catégories B et C.

- divise la catégorie "Basse Entropie" en 3 sous-catégories D, E et F.

- n'affecte pas la catégorie "Haute Entropie Relative", contenant les personnes les plus fiables en termes de performances (Ensemble F Figure 5.9), dont les signatures sont complexes, stables, cursives et robustes aux attaques.

5.5 Evaluation des performances par catégorie de personnes

Jusqu'à présent, nous avons généré automatiquement des catégories de personnes basées sur l'Entropie Relative indépendamment de tout classifieur. Dans cette partie, nous cherchons à étudier le comportement de ces nouvelles catégories en termes de performances en évaluant des systèmes de vérification sur chaque catégorie de personnes. En même temps, nous confronterons ces performances à celles obtenues sur les catégories de personnes générées avec l'Entropie Personnelle.

Dans cette perspective, comme nous travaillons sur des signatures représentées par les coordonnées seulement, nous exploitons alors pour l'étude des performances uniquement le classifieur basé sur la distance élastique (DTW) [96], réputé pour être efficace dans ce contexte, comme nous l'avons précédemment évoqué dans le chapitre 2 à la Section 2.5.3.

L'étude des performance est effectuée sur les deux bases MCYT-100 [91] et Philips [23, 24], sur chaque catégorie de personnes générée avec l'Entropie Relative Personnelle et sur chaque catégorie de personnes obtenue précédemment avec l'Entropie Personnelle.

5.5.1 Description du classifieur et protocole d'évaluation

Pour l'évaluation des performances, le protocole d'expérimentation que nous avons adopté est comme suit : 20 tirages aléatoires sont réalisés sur les signatures authentiques et les imitations. Chaque tirage contient 5 signatures authentiques

utilisées comme signatures de référence. Pour le test, pour chaque personne de la base MCYT-100, nous utilisons ses 20 signatures authentiques restantes et ses 25 bonnes imitations ; pour chaque personne de la base Philips, nous utilisons ses 25 signatures authentiques restantes, ses 25 imitations statiques (HI), et ses 25 imitations dynamiques (SH). Les taux de fausses acceptations et de faux rejets sont calculés sur l'ensemble des 20 tirages aléatoires.

Le classifieur basé sur la distance élastique utilisé dans cette expérimentation est le même que celui utilisé pour l'évaluation des performances dans le Chapitre 4. Pour rappel, le score de dissimilarité est défini comme comme étant le minimum des 5 distances DTW calculées entre la signature de test et les 5 signatures de référence :

$$Score_{DTW} = D_{min}(test, reference_i), \ i = 1...5 \tag{5.9}$$

5.5.2 Evaluation des performances sur la base MCYT-100

La Figure 5.10 montre les courbes de performance du classifieur DTW sur la base MCYT-100 pour les catégories de personnes générées avec l'Entropie (Figure 5.10.a) et l'Entropie Relative (Figure 5.10.b). Le taux d'erreur au point de fonctionnement *EER* ainsi que les intervalles de confiance à 95% sont donnés dans le Tableau 5.5.

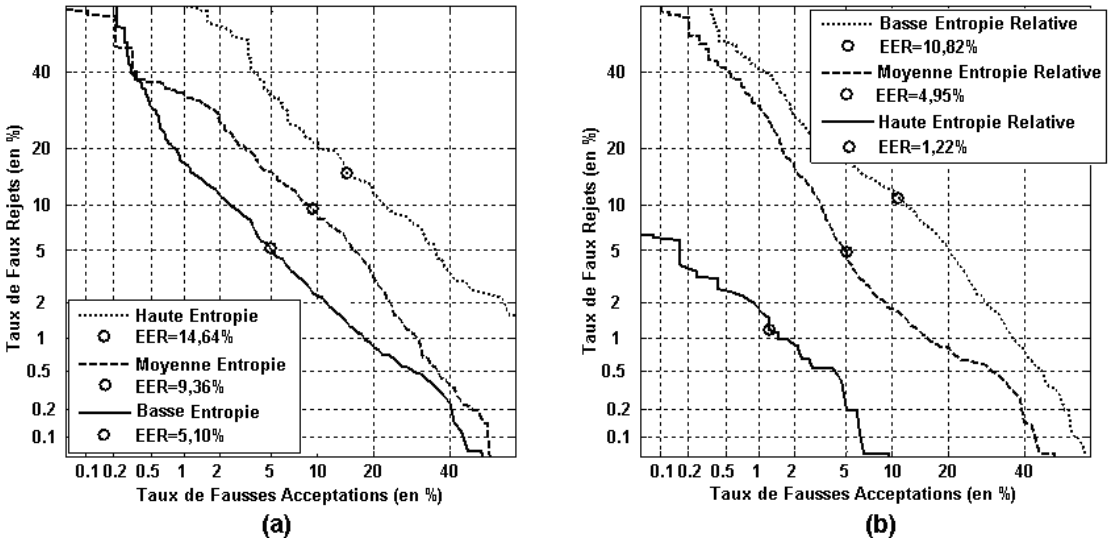

FIGURE 5.10: Courbes de performance sur les catégories de personnes de MCYT-100 générées avec (a) l'Entropie et (b) l'Entropie Relative.

MCYT-100						
Statistiques	Entropie Personnelle			Entropie Relative Personnelle		
	Haute	Moyenne	Basse	Basse	Moyenne	Haute
EER (%)	14,64	9,36	5,10	10,82	4,95	1,22
IC (95%)	$\pm 2,25$	$\pm 0,18$	$\pm 0,15$	$\pm 0,45$	$\pm 0,42$	$\pm 0,40$

TABLEAU 5.5: Taux d'erreur à l'*EER* et intervalles de confiance pour chaque catégorie d'Entropie et d'Entropie Relative de la base MCYT-100.

Au premier abord, on remarque une différence significative dans les performances obtenues sur les diverses catégories de personnes générées soit avec la mesure d'Entropie, soit avec la mesure d'Entropie Relative.

Toutefois, les deux mesures d'entropie conduisent toutes les deux à des résultats cohérents : les signatures qui offrent le plus faible taux d'erreur sont parmi les signatures cursives, les plus complexes et les plus stables.

En l'occurrence, en comparant la meilleure catégorie de personnes en termes de performances pour les deux mesures d'entropie, autrement dit la catégorie "Basse Entropie" et la catégorie "Haute Entropie Relative", on constate que la catégorie "Haute Entropie Relative" est celle qui donne le plus faible taux d'erreur (*EER*=1,22%).

Ainsi, en comparaison avec la catégorie "Basse Entropie", on obtient une amélioration relative des performances de 76,07% à l'*EER* avec la catégorie "Haute Entropie Relative", dont les signatures sont complexes, stables, cursives et les moins vulnérables aux attaques (voir le Tableau 5.5 pour les performances à l'*EER*, et la Figure 5.10 pour les courbes DET). De plus, une telle amélioration se manifeste non seulement à l'*EER*, mais aussi pour tous les points de fonctionnement. En particulier, on remarque sur la Figure 5.10 que l'amélioration est très importante aux faibles valeurs des *FAR*s : un maximum de *FRR* de 6% seulement est atteint avec la catégorie "Haute Entropie Relative" à un *FAR=0*. Par ailleurs, on observe une amélioration importante en termes de *FAR*s. En effet, à un *FRR* égal à 2% par exemple, le taux de *FAR* est 10 fois plus grand sur la catégorie "Basse Entropie" que sur la catégorie "Haute Entropie Relative".

Cette amélioration relative est tout simplement expliquée par le fait que la nouvelle catégorie "Haute Entropie Relative", comme indiqué précédemment dans la Section 5.4.2.1, est un sous-ensemble de la précédente catégorie "Basse Entropie", c'est à dire qu'elle partage les mêmes propriétés que cette dernière, mais en plus elle est plus homogène et ne considère que les signatures les moins

vulnérables aux attaques. Ce qui améliore considérablement les performances sur cette nouvelle catégorie.

Ces résultats prouvent le pouvoir discriminant de notre nouvelle mesure d'Entropie Relative par rapport à l'ancienne mesure d'Entropie pour la catégorisation des personnes.

Par ailleurs, on remarque que les performances s'améliorent aussi sur les deux autres catégories d'Entropie Relative (c.à.d. "Moyenne" et "Basse Entropie Relative") en comparaison avec les catégories "Moyenne" et "Haute Entropie". Cette amélioration est due au fait que ces deux catégories d'Entropie Relative contiennent maintenant un peu plus de signatures qui sont assez stables et complexes contribuant ainsi à diminuer les taux d'erreur sur ces catégories.

Pour un meilleur aperçu concernant l'impact de l'Entropie Relative sur les performances du système de vérification, nous avons tracé sur la Figure 5.11 séparément les courbes de taux de faux rejets (Figure 5.11.a) et de fausses acceptations (Figure 5.11.b) en fonction du seuil de décision. Et ce, sur les 3 premières catégories générées avec l'Entropie Personnelle, ainsi que sur la meilleure catégorie de personnes générée avec l'Entropie Relative Personnelle, celle contenant les signatures stables, complexes, cursives et les moins vulnérables aux attaques.

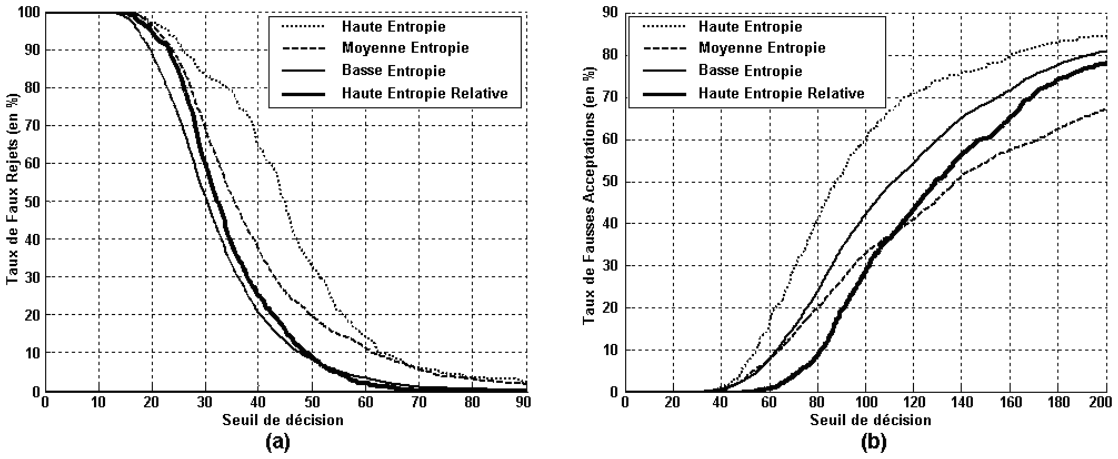

FIGURE 5.11: Courbes de taux de faux rejets (a) et de taux de fausses acceptations (b) en fonction du seuil de décision sur la base MCYT-100, pour les 3 catégories générées avec la mesure d'Entropie ainsi que la catégorie contenant les personnes les plus robustes aux attaques.

La Figure 5.11.b montre que le taux de fausses acceptations *(FAR)* est considérablement faible pour la catégorie "Haute Entropie Relative" en comparaison avec la catégorie "Basse Entropie" : pour les faibles valeurs de FAR, et pour

un seuil de décision entre 50 et 80 par exemple, les valeurs de *FAR* sur la catégorie "Haute Entropie Relative" sont au moins 3,8 fois plus faibles par rapport à la catégorie "Basse Entropie". Effectivement, la catégorie "Haute Entropie Relative" contient des personnes dont leurs signatures authentiques montrent une grande divergence de Kullback-Leibler à leurs imitations, donc les moins vulnérables aux attaques, ce qui réduit considérablement le taux de fausses acceptations sur cette catégorie de personnes.

D'autre part, sur la Figure 5.11.a, on remarque que cette amélioration du taux de fausses acceptations est au prix d'une très faible dégradation du taux de faux rejets. Cet important résultat montre en fait que notre nouvelle mesure d'Entropie Relative Personnelle parvient à faire un bon compromis entre l'amélioration du *FAR* et la dégradation du *FRR*. Du coup, elle est très bien adaptée aux systèmes biométriques confrontés souvent au problème de la discrimination entre les données authentiques et les imitations.

5.5.3 Evaluation des performances sur la base Philips

Les Figures 5.12 et 5.13 montrent les courbes de performance du classifieur DTW sur la base Philips, en considérant respectivement les imitations statiques (HI) et les imitations dynamiques (SH). L'évaluation des performances est effectuée pour les différentes catégories de personnes générées avec l'Entropie (Figures 5.12.a et 5.13.a) et avec l'Entropie Relative (Figures 5.12.b et 5.13.b). Le taux d'erreur au point de fonctionnement *EER* et les intervalles de confiance à 95% sont donnés dans les Tableaux 5.6 et 5.7.

En comparant les courbes de performances sur les Figures 5.12 et 5.13, on remarque que les catégories "Basse Entropie" et "Haute Entropie Relative" sont les meilleures catégories en termes de performances du système de vérification. Aussi, comme observé sur la base MCYT-100, la catégorie "Haute Entropie Relative" est celle qui donne le plus faible taux d'erreur par rapport à la catégorie "Basse Entropie". Et ce, pour les deux types d'imitations. En effet, en comparaison avec la catégorie "Basse Entropie", on obtient une amélioration relative des performances de 100% à l'*EER* avec la catégorie "Haute Entropie Relative" en considérant les imitations statiques (voir le Tableau 5.6) et de 39,11% en considérant les imitations dynamiques (voir le Tableau 5.7).

Ainsi, l'Entropie Relative est clairement un bien meilleur critère que l'Entropie, d'autant plus que la séparation des différentes catégories de personnes en

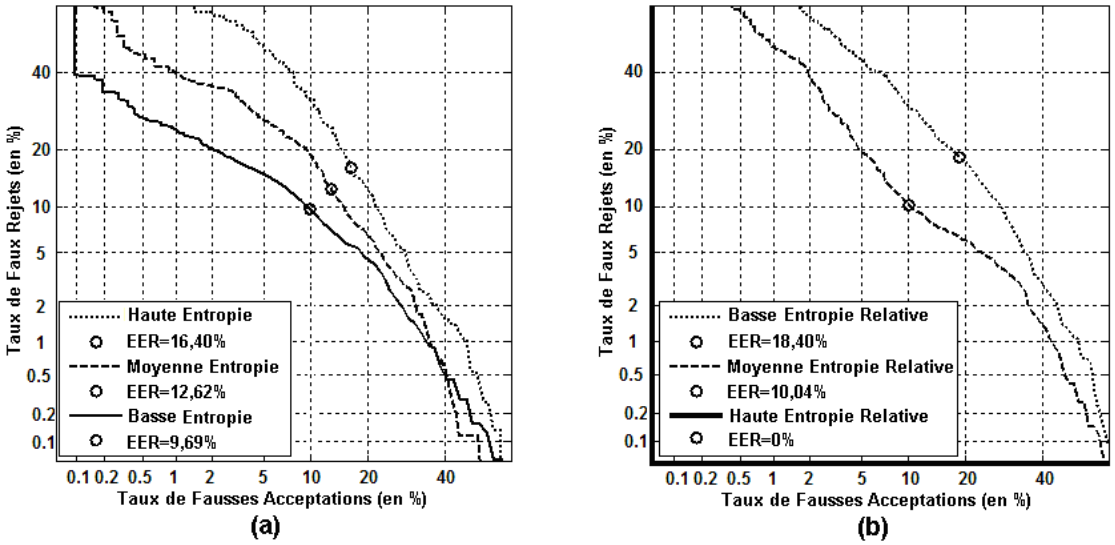

FIGURE 5.12: Courbes de performance sur les catégories de personnes de la base Philips générées avec (a) l'Entropie et (b) l'Entropie Relative en considérant les imitations statiques.

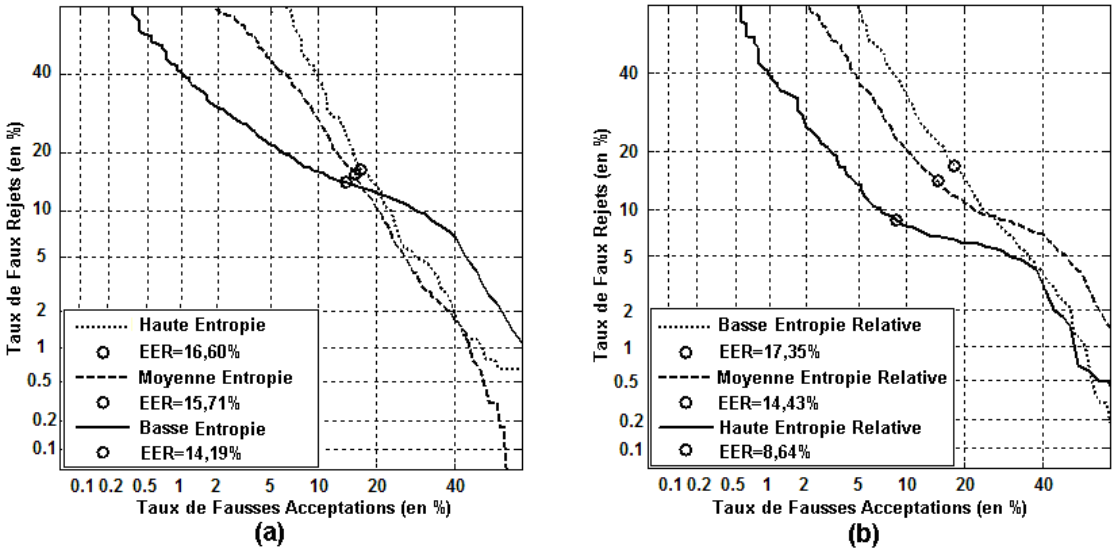

FIGURE 5.13: Courbes de performance sur les catégories de personnes de la base Philips générées avec (a) l'Entropie et (b) l'Entropie Relative en considérant les imitations dynamiques.

Base Philips - imitations statiques (HI)						
Statistiques	Entropie Personnelle			Entropie Relative Personnelle		
	Haute	Moyenne	Basse	Basse	Moyenne	Haute
EER (%)	16,40	12,62	9,69	18,40	10,04	0
IC (95%)	$\pm 3,13$	$\pm 1,51$	$\pm 1,17$	$\pm 1,48$	$\pm 0,58$	± 0

TABLEAU 5.6: Taux d'erreur à l'*EER* et intervalles de confiance pour chaque catégorie d'Entropie et d'Entropie Relative de la base Philips en considérant uniquement les imitations statiques.

Philips - imitations dynamiques (SH)						
Statistiques	Entropie Personnelle			Entropie Relative Personnelle		
	Haute	Moyenne	Basse	Basse	Moyenne	Haute
EER (%)	16,60	15,71	14,19	17,35	14,43	8,64
IC (95%)	$\pm 3,52$	$\pm 1,25$	$\pm 0,96$	$\pm 0,92$	$\pm 1,09$	$\pm 0,59$

TABLEAU 5.7: Taux d'erreur à l'*EER* et intervalles de confiance pour chaque catégorie d'Entropie et d'Entropie Relative de la base Philips en considérant uniquement les imitations dynamiques.

termes de performances est beaucoup plus forte et accentée avec le critère d'Entropie Relative (voir les Tableaux 5.6 et 5.7). D'ailleurs, en considérant les imitations dynamiques, on remarque sur la Figure 5.13 que l'Entropie Personnelle ne fonctionne pas bien et a atteint ses limites avec ce type d'imitations : pour les grandes valeurs de *FAR*s sur la Figure 5.13.a, on remarque que la catégorie "Basse Entropie" change de comportement et devient la pire catégorie en termes de performances ; en revanche, on constate sur la Figure 5.13.b que la catégorie "Haute Entropie Relative" est la meilleure catégorie en termes de performances sur tous les points de fonctionnement de la courbe.

En comparant à présent les performances de la catégorie "Haute Entropie Relative" avec les deux types d'imitations, les Figures 5.12.b et 5.13.b montrent que les performances sur cette catégorie diffèrent selon le type d'imitation : un taux d'erreur *EER* de 0% est atteint en considérant les imitations statiques contre un taux d'erreur *EER* de 8,64% sur les imitations dynamiques (se référer aux Tableaux 5.6 et 5.7). Cette différence significative dans les performances du classifieur DTW sur les deux types d'imitations est cohérente avec notre analyse dans la Section 5.4.1.2 : les imitations dynamiques sont de meilleure qualité que les imitations statiques, comme démontré par la nette différence dans les valeurs d'Entropie Relative, lorsque cette dernière est calculée sur les deux types d'imitations (les faibles valeurs de la divergence de Kullback-Leibler sont obtenues entre les signatures authentiques et les imitations "dynamiques").

Pour un meilleur aperçu concernant l'impact de l'Entropie Relative sur les performances du système de vérification sur la base Philips, nous avons tracé séparément, sur les Figures 5.14 et 5.15, les courbes de taux de faux rejets et de fausses acceptations en fonction du seuil de décision, en considérant respectivement les imitations statiques et dynamiques. Et ce, sur les 3 premières catégories générées avec l'Entropie, ainsi que sur la meilleure catégorie de personnes générée avec l'Entropie Relative ("Haute Entropie Relative").

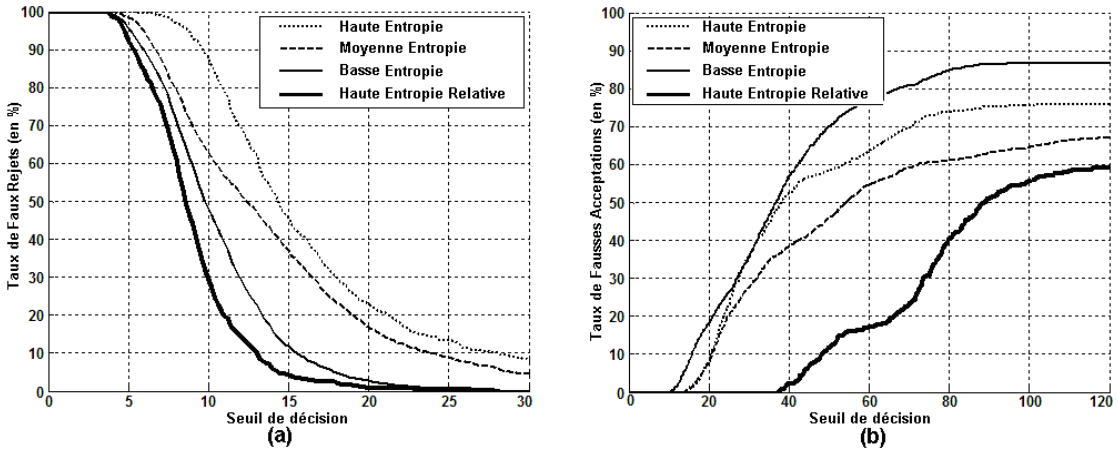

FIGURE 5.14: Courbes de taux de faux rejets (a) et de fausses acceptations (b) en fonction du seuil de décision sur la base Philips, en considérant les imitations statiques, pour les 3 catégories générées avec la mesure d'Entropie ainsi que la catégorie contenant les personnes les plus robustes aux imitations statiques.

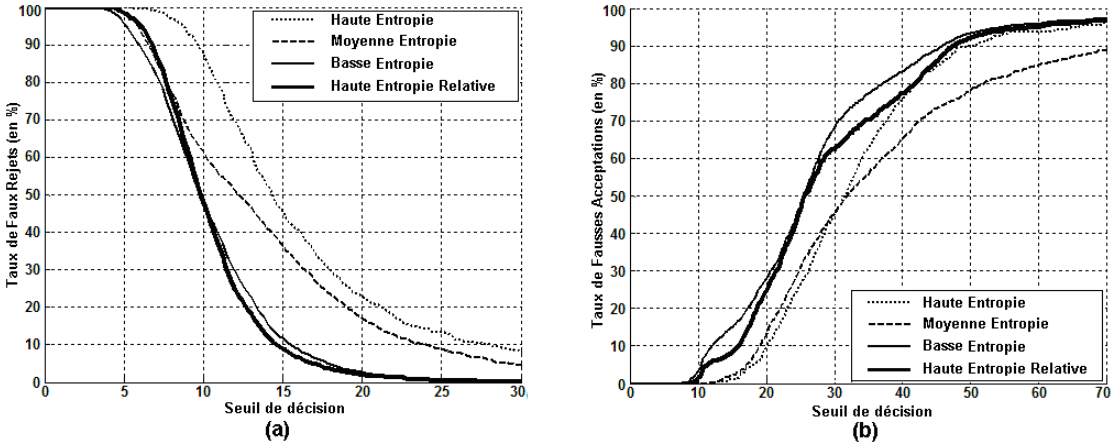

FIGURE 5.15: Courbes de taux de faux rejets (a) et de fausses acceptations (b) en fonction du seuil de décision sur la base Philips, en considérant les imitations dynamiques, pour les 3 catégories générées avec la mesure d'Entropie ainsi que la catégorie contenant les personnes les plus robustes aux imitations dynamiques.

En ce qui concerne les imitations statiques, la Figure 5.14 montre que les valeurs les plus faibles de *FAR* et de *FRR* sont obtenues sur la catégorie "Haute Entropie Relative", celle contenant les signatures stables, complexes, cursives et les moins vulnérables aux attaques. Aussi, on remarque que sur cette catégorie de personnes, pour des valeurs de seuil entre 30 et 38, les taux de fausses acceptations et de faux rejets sont tous les deux à 0%. Par conséquent, il n'y a pas de chevauchement entre les distributions des scores DTW des signatures authentiques et des imitations pour cette catégorie de personnes dans le cas d'imitations statiques. Cela conduit à un taux d'erreur *EER* de 0% (comme montré sur la Figure 5.12.b).

En ce qui concerne les imitations dynamiques, on remarque sur la Figure 5.15

que les *FARs* et*FRRs* sont plus faibles sur la catégorie "Haute Entropie Relative" que sur la catégorie "Basse Entropie". Cependant, dans ce cas bien précis, le comportement des catégories "Haute" et "Moyenne" Entropie attire notre attention : les plus faibles valeurs de *FAR* sont obtenues sur ces deux catégories, et non pas sur la catégorie "Haute Entropie Relative" comme nous avons eu l'habitude de voir jusqu'à présent. Néanmoins, le principal résultat reste maintenu : en termes de compromis entre les deux types d'erreur, faux rejets et fausses acceptations, la catégorie "Haute Entropie Relative" reste de loin la meilleure (bien visible avec les courbes DET sur la Figure 5.13).

5.6 Catégories de personnes d'Entropie et d'Entropie Relative vs. "Ménagerie Biométrique"

Dans le Chapitre 4, nous avons comparé les catégories générées avec la mesure d'Entropie Personnelle à celles de la "Ménagerie Biométrique". Nous avons trouvé que la catégorie "Haute Entropie" correspond aux "Goats" (personnes difficiles à reconnaître) et que les catégories "Moyenne" et "Basse Entropie" correspondent à la catégorie des "Sheeps" (personnes faciles à reconnaître). Cependant, l'Entropie Personnelle n'était pas suffisante pour aller plus loin dans la classification, car elle ne tient compte que des données authentiques à l'inverse de la "Ménagerie Biométrique" qui tient compte des données authentiques ainsi que des imitations. Comme nous l'avons vu dans ce chapitre, l'Entropie Relative Personnelle quantifie la vulnérabilité des signatures aux attaques et vient alors compléter la catégorisation des personnes générée avec l'Entropie Personnelle et permet ainsi une comparaison complète avec la "Ménagerie Biométrique".

Cependant, pour comparer la "Ménagerie Biométrique" à notre catégorisation, nous n'allons pas utiliser uniquement l'Entropie Relative, mais la combinaison de l'Entropie et l'Entropie relative. En effet, bien que l'Entropie Relative soit plus complète que l'Entropie et plus adaptée aux problèmes biométriques, nous avons vu dans le paragraphe 5.4.1.1 que la corrélation entre les deux mesures d'Entropie n'est pas très forte (-0,45). Ainsi, l'utilisation des deux mesures d'entropie permet une catégorisation de personnes plus fine en termes à la fois de leurs données authentiques et de leurs imitations, comme montré dans la Section 5.4.3. par la catégorisation 2D des personnes de MCYT-100.

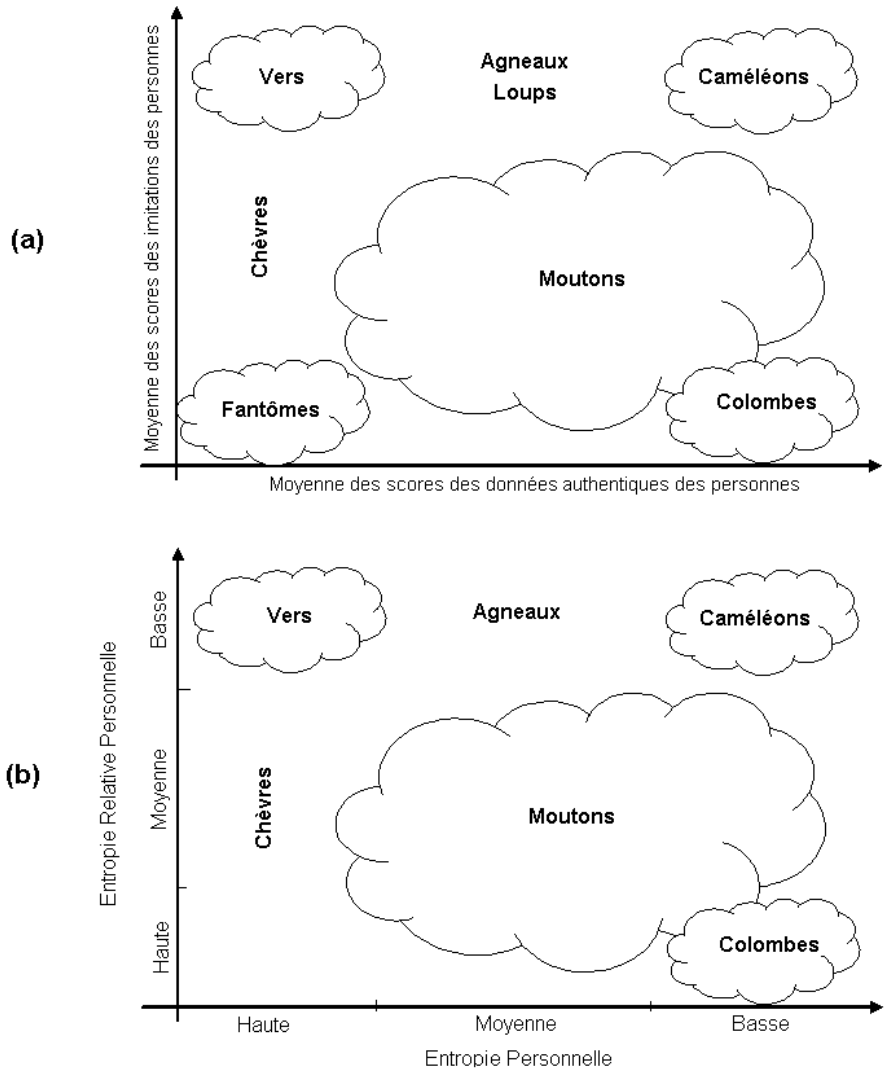

FIGURE 5.16: Comparaison entre la catégorisation de personnes issue (a) de la "Ménagerie Biométrique" et (b) des deux mesures d'Entropie.

La Figure 5.16 présente la classification de la "Ménagerie Biométrique" (Figure 5.16.a) et celle issue de la combinaison d'Entropie et d'Entropie Relative (Figure 5.16.b).

On remarque que les catégories des *Moutons* (*Sheeps*) et des *Chèvres* (*Goats*) sont identifiées par la mesure d'Entropie Personnelle uniquement. Alors que les Agneaux (*Lambs*) sont identifiés par la mesure d'Entropie Relative Personnelle uniquement. Quant aux catégories des Colombes (*Lambs*), Vers (Worms), Caméléons (Chameleons) et Fantômes (Phantoms) elles sont identifiées par la combinaison des deux mesures d'entropie simultanément.

Hormis la catégorie des *Loups* (*wolves*) et des *Fantômes* (*Phantoms*), on remarque que toutes les autres catégories de la "Ménagerie Biométrique" sont

identifiées par nos deux mesures d'Entropie combinées.

En fait, les *Loups* (*wolves*) ne sont pas identifiés par nos mesures d'Entropie car ils correspondent aux bons imitateurs et nous n'avons pas caractérisé les imposteurs dans notre travail.

Quant aux Fantômes(*Phantoms*), nous n'avons pas trouvé ce type de personnes pour les deux bases de signatures MCYT-100 et Philips. Ces personnes présentent les plus fortes valeurs d'Entropie Personnelle (les plus difficiles à caractériser) et les plus fortes valeurs d'Entropie Relative Personnelle (les plus difficiles à imiter). Cela veut dire qu'elles sont très peu complexes et très variables. Ainsi, ces signatures très peu complexes sont faciles à imiter et présentent donc de faibles valeurs d'Entropie Relative. Ce qui est contradictoire avec les caractéristiques de la classe Fantômes. Le seul cas où de telles signatures sont difficiles à imiter est lorsqu'elles présentent une très forte variabilité. Mais dans ce cas, il ne s'agirait pas simplement d'une forte variabilité mais de présence d'anomalies dans les signatures de la personne. En effet, comme les signatures sont très peu complexes et presque toutes différentes, l'imposteur n'arrive pas à imiter la signature de cette personne. Et ce type de phénomène n'a pas été observé dans les bases MCYT-100 et Philips.

5.7 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle mesure de qualité personnelle, dénommée "Entropie Relative Personnelle", qui permet aux personnes d'être caractérisées non seulement en termes de complexité, de variabilité et de lisibilité de leur signature, comme avec notre mesure d'Entropie Personnelle décrite dans le Chapitre 4, mais aussi en termes de sa vulnérabilité aux attaques.

L'Entropie Relative Personnelle mesure pour une personne donnée la divergence de Kullback-Leibler entre les lois de probabilités locales de ses signatures authentiques et des bonnes imitations qui lui sont associées. Ainsi, plus la divergence est faible, plus les imitations sont considérées proches des signatures authentiques, et plus la personne est vulnérable aux attaques ; plus la divergence est grande, plus la personne est considérée robuste aux attaques.

Après avoir noté une corrélation inverse entre notre nouvelle mesure d'Entropie Relative et notre précédente mesure d'Entropie, nous avons catégorisé les

personnes des bases MCYT-100 et Philips suivant leur valeur d'Entropie Relative. Nous avons montré que cette nouvelle catégorisation permet d'extraire la meilleure catégorie de personnes en termes de performances, la catégorie "Haute Entropie Relative" dont les signatures sont stables, complexes, cursives et les moins vulnérables aux attaques.

Ensuite nous avons confronté notre nouvelle mesure d'Entropie Relative aux performances d'un système de vérification, en considérant plusieurs types d'imitations. Et nous avons comparé ces performances à celles précédemment obtenues avec la mesure d'Entropie. Les résultats ont montré que, par rapport à l'ancienne catégorie "Basse Entropie", les performances du classifieur DTW sont nettement meilleures sur la nouvelle catégorie "Haute Entropie Relative" : l'amélioration relative est de 76.07% sur la base MCYT-100 ; sur la base Philips, l'amélioration est de 100% avec les imitations statiques et de 39,11% avec les imitations dynamiques.

Par ailleurs, en analysant les performances du classifieur en termes de fausses acceptations et de faux rejet séparément, nous avons démontré que le taux de fausses acceptations est nettement inférieur sur la catégorie "Haute Entropie Relative" que sur la catégorie "Basse Entropie", au prix quelquefois d'une légère dégradation du FRR ; en d'autres termes, un bon compromis a été observé entre l'amélioration du FAR et la dégradation du FRR sur la catégorie "Haute Entropie Relative".

Lorsque l'on compare nos résultats à ceux de la littérature, il est important de noter, qu'au lieu de mesurer la difficulté d'une signature à être imitée par un imposteur en utilisant uniquement les signatures authentiques de la personne, comme effectué par *Brault et al.* dans [15], nous avons proposé de mesurer la vulnérabilité des signatures aux attaques en tenant compte également des imitations. Aussi, nos résultats convergent avec les conclusions de *Ballard et al.* [6] dans le contexte de l'écriture manuscrite en-ligne : les écritures cursives sont les moins vulnérables aux attaques. En effet, comme montré sur la base MCYT-100, la catégorie "Haute Entropie Relative", comportant les signatures les plus robustes aux attaques, contient des signatures se rapprochant parfois de l'écriture cursive. Cependant, toutes les signatures cursives ne sont pas robustes aux attaques. En effet, nous avons trouvé avec notre mesure d'Entropie Relative que certaines signatures ayant l'aspect d'écriture cursive peuvent être vulnérables aux attaques. Ainsi, notre mesure d'Entropie Relative arrive à détecter ces signatures de manière quantitative et automatique, ce qui n'est pas possible dans [6] puisque la catégorisation

de personnes est effectuée par des experts humains qui ont tendance à caractériser les signatures de manière grossière.

En conclusion, notre approche présente deux avantages principaux : d'abord, au lieu de classer les styles d'écriture par un expert humain comme dans [6], la catégorisation des personnes avec la mesure d'Entropie Relative est automatique ; d'autre part, au lieu d'utiliser un classifieur pour caractériser la vulnérabilité des personnes aux attaques, cette vulnérabilité est quantifiée directement par le biais de notre mesure d'Entropie Relative, indépendamment de tout classifieur.

Chapitre 6

Mesure de qualité des imitations

En général, un système de vérification de signature en-ligne est évalué sur des bases de signatures contenant à la fois des signatures authentiques et des imitations. Nous avons vu au Chapitre 2 qu'il existe essentiellement deux types d'imitations : les imitations aléatoires et les bonnes imitations. Ces dernières peuvent être statiques, dynamiques ou professionnelles.

Outre le type des imitations, il y a aussi le concept de la "qualité" des imitations. Cette qualité peut être très variable pour plusieurs raisons : d'une part, un imposteur peut être plus habile qu'un autre à imiter une signature authentique ; d'autre part, en imitant une signature, l'imposteur n'est pas aussi confortable qu'il le serait lors de la réalisation de sa propre signature, tout simplement parce que ce n'est pas son geste naturel. Cette constatation suggère qu'il existe une qualité intrinsèque à chaque signature imitée qui doit être quantifiée de façon indépendante de celui qui la fait.

Nous proposons dans ce chapitre une mesure permettant de quantifier la qualité des imitations indépendamment de toute information sur la qualité ou la nature de la signature "vraie" qui est imitée, ce qui est totalement nouveau dans la littérature ; son impact est important car elle permet lors de l'évaluation de systèmes de vérification, de catégoriser les signatures imitées et d'évaluer les performances du système de reconnaissance en fonction de la qualité des imitations, et permet aussi d'apporter une information vis-à-vis de la résistance des systèmes aux attaques.

Dans ce chapitre, nous commencerons par présenter notre nouvelle mesure de qualité des imitations, qui s'appuie sur la mesure d'Entropie Personnelle introduite dans le Chapitre 4. Puis, nous la confronterons aux performances de notre

classifieur basé sur la distance élastique, sur la base de signature MCYT-100, et sur la base Philips avec ses différents types d'imitations disponibles. Ensuite, sur les deux bases BioSecure DS2 et DS3, nous étudierons l'impact des conditions et du protocole d'acquisition sur la qualité des imitations. Aussi, nous comparerons nos résultats avec ceux de la littérature, et montrerons l'efficacité de notre mesure de qualité pour quantifier la qualité des différents types d'imitations.

6.1 Mesure de qualité des imitations

De manière analogue à nos précédents travaux présentés dans les Chapitres 4 et 5, les signatures sont décrites dans ce chapitre uniquement par les coordonnées *(x,y)*. Ce choix est motivé essentiellement par le fait que les coordonnées du stylet sont les seules fonctions temporelles communes à toutes les bases de signatures existantes, qu'elles soient acquises sur une tablette graphique ou sur un PDA. Ainsi dans ce chapitre, nous considérons 4 bases de signatures en-ligne, dont l'une contient des signatures acquises sur un PDA, pour laquelle seules les coordonnées *(x,y)* sont disponibles.

6.1.1 Rappel sur la mesure d'Entropie Personnelle

Comme détaillé dans la Section 4.3.2, l'Entropie Personnelle est mesurée au moyen d'une estimation locale des densités de probabilité après apprentissage d'un Modèle de Markov Caché (MMC) sur 10 signatures authentiques d'une personne donnée [43] . En d'autres termes, une variable aléatoire Z_i est associée à chaque portion fixe S_i de la signature, générée par l'algorithme de Viterbi [96] en fonction du MMC associé à la personne, et l'entropie de cette portion $H(Z_i)$ est calculée sur l'ensemble des réalisations de la variable Z_i, comme suit :

$$H(Z_i) = -\sum_{z \in Z_i} p(z)\, log_2\, p(z) \tag{6.1}$$

Ensuite, l'entropie d'une signature authentique *"sig"* est calculée en moyennant les valeurs d'entropie locales $H_{sig}(Z_i)$ sur l'ensemble des portions S_i, puis normalisée par la durée de la signature :

$$H_{sig}(Z) = \frac{1}{N * T} \sum_{i=1}^{N} H_{sig}(Z_i) \tag{6.2}$$

où T est la durée de la signature en secondes et N le nombre de portions générées par le MMC (algorithme de Viterbi). Ainsi, nous obtenons une mesure d'Entropie exprimée en bits par seconde.

Finalement, cette mesure est moyennée sur les 10 signatures authentiques considérées pour obtenir une mesure d'entropie par personne H_p, dénommée "Entropie Personnelle" [43].

Jusqu'à présent, nous avons simplement rappelé brièvement le calcul de la notion d'Entropie Personnelle. Dans ce qui suit, nous détaillerons comment ce concept est utilisé pour mesurer la qualité d'une imitation.

6.1.2 Mesure de qualité d'une imitation avec l'Entropie Personnelle

Une imitation est évidemment liée au tracé de la signature authentique cible. Nous avons montré qu'il est effectivement possible de caractériser localement la fonction de densité de probabilité de chaque portion de la signature au moyen d'un MMC [43]. De cette manière, après l'étape de segmentation, nous avons à notre disposition, via ce MMC, la loi de probabilité locale (sur chaque portion) régissant la réalisation de la signature authentique cible.

Pour mesurer la qualité d'une imitation, l'idée est de tirer partie de l'information statistique relative à la signature authentique cible pour mesurer, comme qualité d'une imitation, combien cette dernière va coïncider à la fonction de densité de probabilité "idéale", celle de la personne cible. Cependant, comme mentionné ci-dessus, les lois de probabilité locales régissant la réalisation de la signature authentique cible sont estimées par le MMC utilisé pour calculer l'Entropie Personnelle associée à la personne cible. Par conséquent, nous proposons de quantifier la qualité d'une imitation Q_{im} comme étant la dissimilarité existant entre l'Entropie Personnelle H_p propre à la personne cible, et l'entropie de l'imitation $H_{im}(Z)$ calculée avec le même MMC, à savoir celui de cette personne cible :

$$Q_{im} = |H_{im}(Z) - H_p| \tag{6.3}$$

Plus précisément, $H_{im}(Z)$ est calculée de la même manière que $H_{sig}(Z)$ suivant l'équation 6.2, où Z correspond maintenant à l'ensemble des couples (x, y) appartenant à l'imitation "im".

De ce fait, les imitations montrant des valeurs faibles de cette distance sont jugées proches des signatures authentiques cibles, et ainsi considérées de “bonne” qualité. A l’opposé, les imitations montrant des valeurs élevées de cette distance sont considérées de “mauvaise” qualité.

6.2 Impact de la qualité des imitations sur les performances des systèmes de vérification

Dans cette section, nous cherchons à étudier le comportement de notre nouvelle mesure de qualité des imitations en termes de performances de notre système de vérification, basé sur la distance élastique [96]. Et ce, sur les quatre bases de signatures en-ligne : MCYT-100 [91], Philips [23, 24], et les deux sous-ensembles de la base BioSecure DS2 et DS3 [54, 90] contenant les signatures des mêmes 120 personnes. Ces quatre bases de signatures nous permettent de fournir une preuve expérimentale de la pertinence de notre nouvelle mesure de qualité lorsqu’elle est confrontée à différents protocoles d’acquisition d’imitations et différents types d’imitations.

Dans cette perspective, afin d’analyser les performances de notre système de vérification en fonction de la qualité des imitations, nous commencerons tout d’abord par classer les imitations associées à chaque personne en fonction de leur mesure de qualité, en nous servant de la classification hiérarchique ascendante [89]. Il faut noter que pour chaque personne nous ne disposons au maximum que de 30 signatures imitées. Puis, nous partagerons notre travail en deux parties. Dans la première partie (Section 6.2.2), nous adopterons une catégorisation grossière et préliminaire des imitations : seulement deux catégories d’imitations sont générées par personne, à savoir imitations de “bonne” qualité et imitations de “mauvaise” qualité. Dans la deuxième partie (Section 6.2.3), nous adopterons une catégorisation d’imitations personnalisée en étudiant le nombre optimal de catégories d’imitations par personne pour une analyse plus fine des performances.

6.2.1 Description du classifieur et protocole d’évaluation

Pour l’étude des performances, nous utilisons un classifieur basé sur la distance élastique (DTW) [96], réputé pour être efficace depuis SVC’2004 [122] pour

l'évaluation des systèmes de vérification sur des signatures décrites uniquement par les coordonnées [71, 122, 82].

Le protocole d'expérimentation que nous avons adopté pour l'évaluation des performances est comme suit : 20 tirages aléatoires sont réalisés sur les signatures authentiques et les imitations. Chaque tirage contient 5 signatures authentiques utilisées comme signatures de référence. Pour le test, on considère pour chaque personne uniquement les imitations qui lui sont associées et on ne considère pas des tests avec des signatures authentiques. De ce fait, les performances du système sont évaluées uniquement en termes de taux de fausses acceptations calculés sur l'ensemble des 20 tirages aléatoires.

En effet, notre étude dans ce chapitre se concentre exclusivement sur la mesure de qualité des imitations. Ces dernières ont un impact en termes de performances uniquement sur le taux de fausses acceptations (FAR). Pour cette raison, lorsque nous confrontons notre mesure de qualité aux performances du classifieur DTW, nous le faisons uniquement en termes de FAR.

Le classifieur basé sur la distance élastique utilisé dans ce Chapitre est le même que celui utilisé pour l'évaluation des performances dans les Chapitres 4 et 5. Pour rappel, le score de dissimilarité est défini comme suit :

$$Score_{DTW} = D_{min}(test, reference_i), \;\; i = 1...5 \tag{6.4}$$

6.2.2 Catégorisation préliminaire des imitations

6.2.2.1 Mesure de qualité des imitations confrontée aux imitations statiques de la base MCYT-100

La Figure 6.1 montre un exemple d'une signature authentique de la base MCYT-100 et différentes imitations qui lui sont associées, classées suivant notre mesure de qualité en imitations de "bonne" et "mauvaise" qualité suivant la classification hiérarchique [89]. En se basant uniquement sur l'aspect visuel des signatures, on remarque l'efficacité de notre mesure de qualité : en effet, le tracé des imitations de "bonne" qualité est très proche de celui de la signature authentique en comparaison avec les imitations de "mauvaise" qualité.

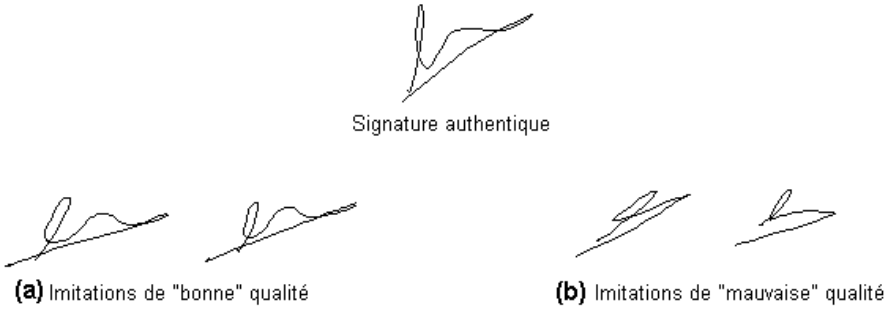

FIGURE 6.1: Exemples d'une signature authentique de la base MCYT-100 et des imitations qui lui sont associées (a) de "bonne" qualité et (b) de 'mauvaise" qualité.

Le Tableau 6.1 présente la distribution des imitations statiques de la base MCYT-100, toutes personnes confondues, après avoir défini le nombre de catégories des imitations à 2 par personne.

Base MCYT-100	Bonnes imitations (Total : 2500)	
	Bonne qualité	Mauvaise qualité
Nombre d'imitations	1433	1067
Pourcentage d'imitations	57,32%	42,68%

TABLEAU 6.1: Distribution des imitations statiques de MCYT-100 en imitations de "bonne" et "mauvaise" qualité.

On remarque que dans la base MCYT-100 environ les deux tiers des imitations disponibles sont de "bonne" qualité (57,32%). Ce résultat confirme les affirmations dans [91], à savoir que des imitations hautement qualifiées (*"highly-shaped"*, en anglais) sont disponibles dans la base MCYT-100.

La Figure 6.2 montre les courbes de taux de fausses acceptations (*FAR*) en fonction du seuil de décision sur la base MCYT-100, pour les deux catégories d'imitations séparément : imitations de "bonne" et de "mauvaise" qualité.

La Figure 6.2 montre l'efficacité de notre mesure de qualité d'imitations : en effet, le *FAR* est beaucoup plus élevé sur la catégorie contenant les imitations de "bonne" qualité que sur celle contenant les imitations de "mauvaise" qualité. On note que pour une valeur du seuil de décision égale à 100, il existe un facteur de 1,53 entre les *FAR*s des deux catégories. En outre, on constate que la pente de la courbe *FAR* associée aux imitations de "bonne" qualité est beaucoup plus forte que celle associée aux imitations de "mauvaise" qualité, qui évolue de façon presque linéaire pour les valeurs de seuil de décision entre 100 et 200.

Dans la section suivante, nous allons étendre notre analyse à la base Philips [23, 24], afin d'étudier le comportement de notre nouvelle mesure de qualité des

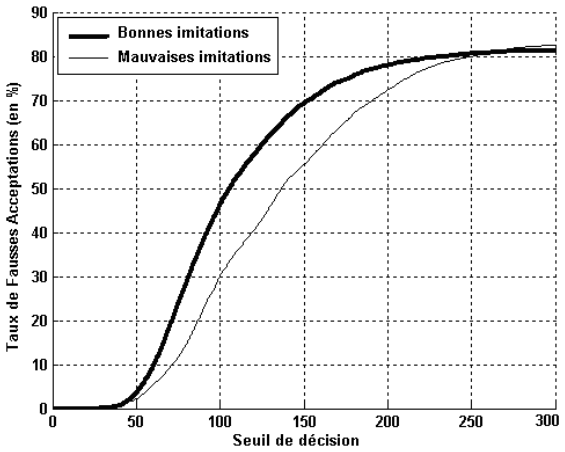

FIGURE 6.2: Courbes de taux de fausses acceptations en fonction du seuil de décision sur les imitations de "bonne" et de "mauvaise" qualité de la base MCYT-100.

imitations en présence de différents types d'imitations : statiques, dynamiques et professionnelles. Nous montrerons ainsi la sensibilité de notre nouvelle mesure de qualité à la nature même de l'imitation à laquelle elle se trouve confrontée.

6.2.2.2 Mesure de qualité des imitations confrontée aux différents types d'imitations de la base Philips

Le Tableau 6.2 montre la distribution des imitations de la base Philips, toutes personnes confondues, après avoir fixé à 2 le nombre de catégories des imitations. Et ce, indépendamment du type des imitations disponibles dans la base Philips : les imitations statiques, dynamiques et professionnelles sont mélangées ensemble pour cette catégorisation préliminaire des imitations.

On constate que dans la base Philips près de trois-cinquièmes des imitations disponibles sont de "bonne" qualité (58,82%).

Base Philips	Bonnes imitations (Total : 3140)	
	Bonne qualité	Mauvaise qualité
Nombre d'imitations	1847	1293
Pourcentage d'imitations	58,82%	41,18%

TABLEAU 6.2: Distribution des imitations de la base Philips en imitations de "bonne" et de "mauvaise" qualité.

Aussi, la Figure 6.3 montre que notre mesure de qualité se comporte bien : le *FAR* est beaucoup plus élevé sur la catégorie contenant les imitations de "bonne" qualité que sur celle contenant les imitations de "mauvaise" qualité. En effet, on

note que pour une valeur du seuil de décision égale à 40, il existe un facteur de 1,54 entre les *FARs* des deux catégories.

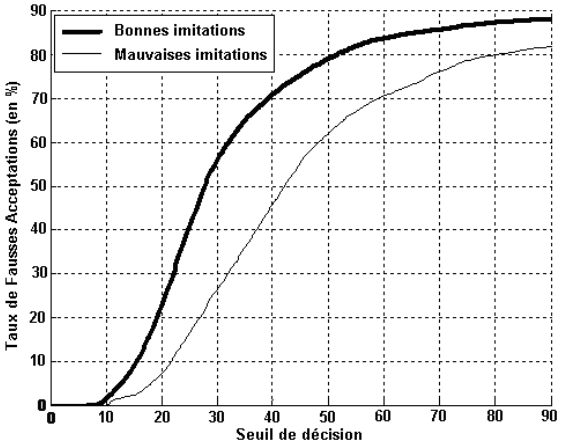

FIGURE 6.3: Courbes de taux de fausses acceptations en fonction du seuil de décision sur les imitations de "bonne" et de "mauvaise" qualité de la base Philips.

Pour une analyse plus fine de notre mesure de qualité des imitations et une meilleure distinction entre les différents types d'imitations présents dans la base Philips, nous avons représenté sur la Figure 6.4 des histogrammes illustrant le pourcentage d'imitations de "bonne" et de "mauvaise" qualité pour chaque type d'imitation.

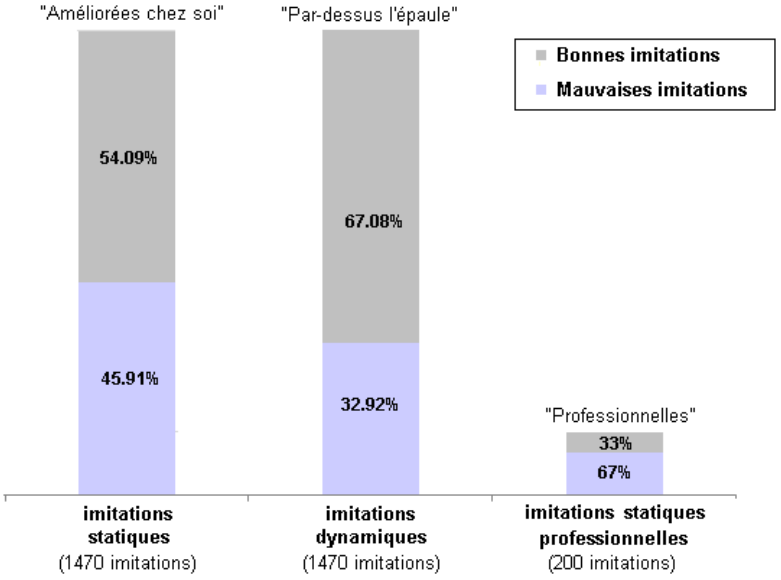

FIGURE 6.4: Pourcentage d'imitations de "bonne" et de "mauvaise" qualité pour chaque type d'imitations de la base Philips : statiques, dynamiques et professionnelles.

On observe que plus du deux-tiers des imitations "par-dessus l'épaule" sont de "bonne" qualité (67,08%), alors que pour les imitations "améliorées chez soi",

la distribution des imitations de “bonne” et de “mauvaise” qualité est presque uniforme. Ce résultat conforte notre connaissance sur le fait que les imitations dynamiques sont de bien meilleure qualité que les imitations statiques dans le cadre en-ligne ; aussi, ce résultat est cohérent avec ce que *Dolfing* a obtenu sur ces deux types d'imitations en se servant d'un classifieur basé sur un Modèle de Markov Caché [23, 24] : les imitations dynamiques sont deux fois plus acceptées par le classifieur que les imitations statiques.

Un autre résultat intéressant émerge : seulement 33% des imitations statiques “professionnelles”, faites par des experts légistes, sont de “bonne” qualité. De plus, on constate que ces imitations sont de moins bonne qualité que les imitations “améliorées chez soi”, qui sont des imitations statiques faites par des amateurs. On constate alors une inversion de tendance par rapport à nos attentes, car les imitateurs professionnels sont expérimentés dans la détection de fraude. Cela peut s'expliquer par le fait que les experts légistes exploitent leur savoir-faire acquis dans le cadre hors-ligne pour faire des imitations dans le cadre en-ligne. Dans le cadre en-ligne, ces experts vont en effet reproduire des imitations selon leur méthode de travail dans le hors-ligne, omettant ainsi l'aspect dynamique de la signature.

D'après ce dernier résultat, nous pouvons conclure que notre mesure de qualité des imitations souligne l'importance de l'aspect dynamique pour la production d'une bonne imitation dans le contexte en-ligne : plus l'imposteur tente de s'en tenir uniquement à la reproduction de la forme de la signature cible, plus la qualité de l'imitation qui en résulte est “mauvaise”. Ceci est en fait affirmé par ce qu'on a observé avec notre mesure de qualité sur les imitations “par-dessus l'épaule” : ces imitations sont de loin celles qui sont de meilleure qualité.

Afin de valider ces résultats fondés sur notre mesure de qualité des imitations, nous les avons comparés à ceux obtenus antérieurement dans la littérature par *Dolfing* [23, 24], qui s'est basé uniquement sur les performances du classifieur MMC. Nous avons constaté que nos résultats aboutissent aux mêmes conclusions que *Dolfing* dans [23, 24] : d'une part, comme noté par *Dolfing*, nous confirmons par le biais de notre mesure de qualité des imitations que les experts légistes seraient de bien meilleurs faussaires dans le cadre hors-ligne que dans celui du en-ligne ; d'autre part, notre mesure de qualité des imitations confirme ce que *Dolfing* a observé dans [23, 24] : il ne parvenait pas à obtenir un plus grand taux d'erreur sur les imitations “professionnelles” par rapport à l'autre type d'imitations statiques faites par des amateurs, à savoir les imitations “améliorées chez soi”. En effet, notre mesure de qualité des imitations montre que les imitations “professionnelles” sont de moins

bonne qualité que les imitations "améliorées chez soi" faites par des amateurs. Du coup, les imitations "professionnelles" sont faciles à détecter par un système de vérification automatique.

Ainsi, notre nouvelle mesure de qualité présente l'avantage de confirmer, de manière quantitative et indépendante de l'étape de classification, ce que *Dolfing* a observé dans la littérature [23, 24] en évaluant les performances de son classifieur MMC sur chaque type d'imitations.

Dans la section suivante, nous continuons l'analyse de notre mesure de qualité des imitations sur les deux sous-ensembles BioSecure DS2-120 et DS3-120 [54, 90], contenant les mêmes 120 personnes, et acquis sur différentes sortes de plate-forme : l'un sur une plateforme fixe (tablette graphique) et l'autre sur une plateforme mobile (PDA). Ce contexte nous permettra d'étudier l'influence du protocole et des conditions d'acquisition sur la qualité des imitations.

6.2.2.3 Impact du protocole et des conditions d'acquisition sur la qualité des imitations dans les bases BioSecure DS2 et DS3

Afin d'étudier l'impact des conditions d'acquisition sur la qualité des imitations, nous avons en premier temps rassemblé pour chaque personne toutes ses imitations de DS2 et DS3 (nous rappelons que DS2 et DS3 contiennent les mêmes 120 personnes). Puis, nous avons classé les imitations de chaque personne (de DS2 et DS3 confondues) en 2 catégories : imitations de "bonne" et "mauvaise" qualité. A noter que le fait de rassembler les imitations de DS2 et DS3 avant de les catégoriser permet de classer les imitations de DS2 et DS3 en adoptant une même séparatrice entre la catégorie d'imitations de "bonne" qualité et celle des imitations de "mauvaise" qualité. Ceci est nécessaire pour pouvoir par la suite comparer les résultats entre DS2 et DS3. Ensuite, nous avons calculé par personne le nombre d'imitations de "bonne" qualité (respectivement de "mauvaise" qualité) appartenant à DS2 et à DS3 séparément.

Les Tableaux 6.3 et 6.4 montrent la distribution des imitations de DS2-120 et DS3-120 séparément, toutes personnes confondues, après avoir défini le nombre de catégories des imitations à 2.

Base DS2-120	Bonnes imitations (Total : 1200)	
	Bonne qualité	Mauvaise qualité
Nombre d'imitations	704	496
Pourcentage d'imitations	58,67%	41,33%

TABLEAU 6.3: Distribution des imitations de la base DS2-120 en imitations de "bonne" ou "mauvaise" qualité.

Base DS3-120	Bonnes imitations (Total : 1200)	
	Bonne qualité	Mauvaise qualité
Nombre d'imitations	766	434
Pourcentage d'imitations (%)	63,83	36,17

TABLEAU 6.4: Distribution des imitations de la base DS3-120 en imitations de "bonne" ou "mauvaise" qualité.

Contrairement à ce qu'on aurait pu prédire, en comparant DS2 et DS3, on remarque étonnamment que le sous-ensemble DS3, dont les signatures sont acquises sur une plateforme mobile, contient 5,16% de plus d'imitations de "bonne" qualité que DS2, dont les signatures sont acquises sur une plateforme fixe.

Cela peut s'expliquer par deux facteurs : d'abord, le protocole d'acquisition des imitations de DS3 est mieux adapté à la capture d'imitations de bonne qualité, puisqu'une interface d'acquisition spécifique fournit à l'imposteur en même temps des informations statiques et dynamiques sur la signature à imiter [90, 54]. En effet, l'imposteur visualise sur l'écran tactile du PDA la séquence d'écriture de la signature cible, puis signe directement sur l'image représentant la trajectoire du stylet [90, 54]. Via cet outil interactif, l'imposteur ne peut que se rapprocher à la fois de la forme et de la dynamique de la signature cible. Le protocole d'acquisition de la base BioSecure DS3 est décrit en détail dans la Section 2.4.4.2.

Cette interface contribue de manière significative à la production d'imitations de très bonne qualité, comme il apparaît clairement sur le comportement du *FAR* sur la Figure 6.5.b correspondant à DS3 par rapport à la Figure 6.5.a correspondant à DS2 : l'augmentation du *FAR* est beaucoup plus rapide sur DS3 et les taux d'erreur atteignent des valeurs beaucoup plus élevées sur DS3 que sur DS2 (75% des *FAR*s sur DS3 contre 50% sur DS2).

La deuxième raison de ce phénomène est liée au type de la plateforme utilisée. Comme nous l'avons évoqué auparavant dans la Section 1.4.1.3, la plateforme mobile (PDA) est loin d'être un support auquel le signataire est familiarisé, car ce dernier est plus habitué à signer sur une grande surface plane et fixe sur laquelle il peut poser sa main. Ainsi, en signant sur un PDA, le signataire verra sa signature se raccourcir, devenir moins complexe et plus variable. En termes d'entropie, signer

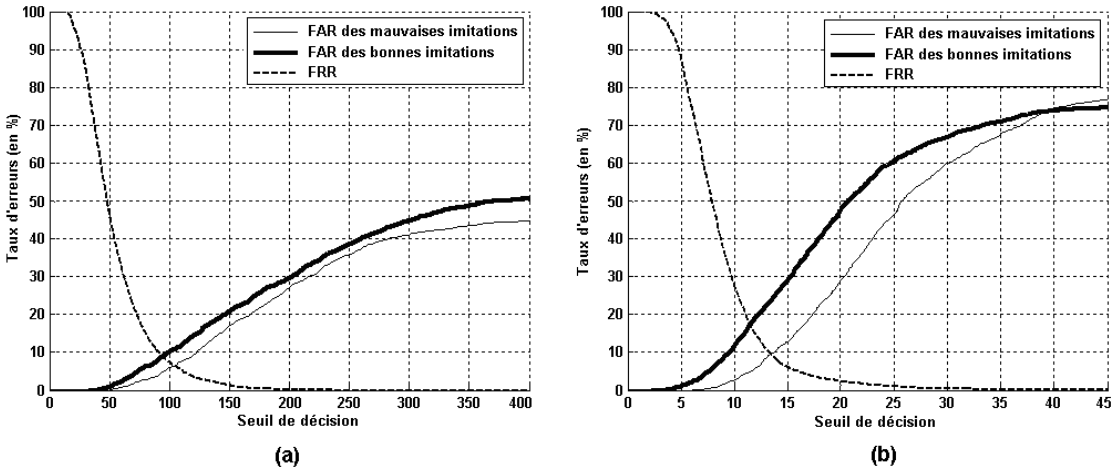

FIGURE 6.5: Courbes de taux de fausses acceptations et de faux rejets en fonction du seuil de décision sur les imitations de "bonne" et de "mauvaise" qualité de (a) DS2-120 et (b) DS3-120 contenant les mêmes personnes.

sur une plateforme mobile induit une augmentation de l'Entropie Personnelle par rapport à une signature réalisée sur une plateforme fixe. Ce fait tend alors à réduire la différence existant entre la valeur d'entropie de l'imitation et l'Entropie Personnelle, suivant l'équation 6.3. Par conséquent, les imitations dans DS3 ont tendance à être "plus proches" des signatures authentiques cibles que dans DS2, et ainsi la qualité de l'imitation dans DS3 augmente car elle dépend d'une certaine manière de la qualité des signatures authentique à imiter.

6.2.3 Catégorisation personnalisée des imitations

Dans la Section 6.2.2, nous avons utilisé une catégorisation préliminaire et grossière des imitations en 2 catégories "extrêmes", à savoir catégories d'imitations de "bonne" et de "mauvaise" qualité. Notre objectif dans cette section est de trouver le nombre approprié de catégories d'imitations pour chaque personne et d'évaluer ensuite son impact sur les performances du classifieur.

Ainsi, en suivant la même procédure décrite en Annexe A, nous faisons appel aux indices de validité permettant de trouver le nombre optimal de catégories d'imitations par personne. Puis, en se basant sur cette nouvelle catégorisation optimale d'imitations, nous évaluerons encore une fois les performances en termes de *FARs* sur les bases MCYT-100 et Philips, et comparerons ces résultats à ceux basés sur notre précédente catégorisation (2 catégories).

6.2.3.1 Evaluation des performances par catégorie d'imitations sur les bases MCYT-100 et Philips

Après avoir étudié le nombre de catégories d'imitations pour chaque personne de MCYT-100 et Philips, nous avons constaté que pour les deux bases de signatures le nombre de catégories d'imitations optimal varie entre 3 et 4 en fonction des personnes. Nous rappelons que pour chaque personne nous ne disposons au maximum que de 30 signatures imitées. Ainsi, dû au faible nombre d'exemples, ces catégories ne seraient peut être pas très stables.

A noter que seulement 13% des personnes dans la base MCYT-100 ont 4 catégories, tandis que 87% ont 3 catégories. Le même phénomène est observé sur la base de signatures Philips, où seulement 23,53% des personnes ont 4 catégories, et 74,45% ont 3 catégories.

Comme peu de personnes ont 4 catégories dans les deux bases de signatures, afin de comparer cette catégorisation personnalisée à l'ancienne catégorisation en 2 catégories (Section 6.1.2), nous considérons pour toutes les personnes 3 catégories d'imitations en fusionnant les 2 catégories intermédiaires pour les personnes ayant 4 catégories d'imitations.

Nous rapportons dans le Tableau 6.5, la nouvelle catégorisation des imitations en 3 catégories pour les deux bases de signatures. Dans le Tableau 6.5 tous les types d'imitations de la base de Philips (statiques, dynamiques et professionnelles) sont mélangés.

Bases	Catégories d'imitations		
	Bonne qualité	Moyenne qualité	Mauvaise qualité
MCYT-100	37,32%	40%	22,68%
Philips	38,44%	38,91%	22,65%

Tableau 6.5: Distribution des imitations des bases MCYT-100 et Philips en "bonne", "moyenne" et "mauvaise" qualité.

Nous remarquons tout d'abord que la répartition des personnes sur les 3 nouvelles catégories est semblable pour les deux bases de signatures. Aussi, un pourcentage élevé d'imitations sont classées comme étant de "bonne" qualité, alors que la proportion des imitations de "mauvaise" qualité est plus faible et par un facteur d'au moins 1,65. Cette classification des imitations confirme que les bases de signatures MCYT-100 et Philips contiennent une proportion acceptable d'imitations de "bonne" qualité. C'est un résultat important car ces bases de signatures

sont largement utilisées dans la littérature pour l'évaluation des performances des systèmes automatiques de vérification de signatures en-ligne.

A présent, afin d'étudier l'impact de notre nouvelle catégorisation des imitations sur les performances de notre système de vérification, nous avons tracé sur la Figure 6.6 les courbes de performance en termes de *FAR* en fonction du seuil de décision, sur les deux bases MCYT-100 et Philips, en considérant seulement 3 catégories d'imitations par personne (comme très peu de personnes ont 4 catégories d'imitations comme expliqué ci-dessus), à savoir imitations de "bonne", "moyenne" et "mauvaise" qualité.

Comme observé avec la catégorisation préliminaire en 2 catégories, la Figure 6.6 montre sur les deux bases de signatures que le taux de fausses acceptations est plus élevé sur la catégorie d'imitations de "bonne" qualité, et est plus faible sur la catégorie d'imitations de "mauvaise" qualité. On remarque aussi, comme prévu, que la courbe de fausses acceptations obtenue sur les imitations de "moyenne" qualité se situe entre celles des deux catégories de qualité extrême.

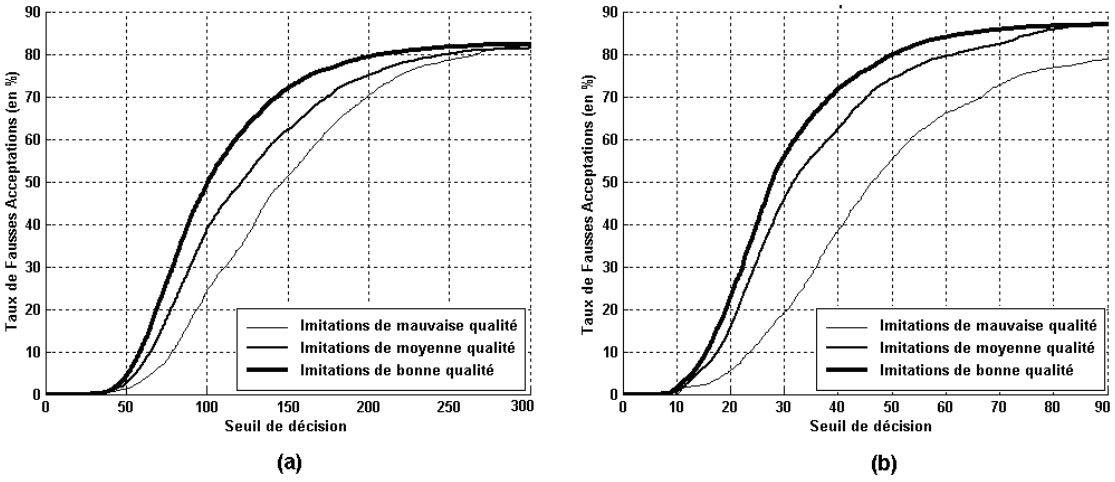

FIGURE 6.6: Courbes de taux de fausses acceptations en fonction du seuil de décision sur les imitations de "bonne", "moyenne" et "mauvaise" qualité des bases (a) MCYT-100 et (b) Philips.

Néanmoins, le principal résultat obtenu avec cette catégorisation d'imitations personnalisée est que la différence entre les *FARs* des catégories de "bonne" et de "mauvaise" qualité s'est considérablement accentuée par rapport à la précédente catégorisation (Figures 6.2 et 6.3). Effectivement, la recherche du nombre optimal de catégories d'imitations par personne conduit à une catégorisation fine des imitations qui accentue encore plus la différence entre les catégories extrêmes. En comparaison avec la catégorisation préliminaire, la séparation relative entre les deux catégories extrêmes s'est améliorée avec la catégorisation personnalisée de

66,67% à un seuil de 100 pour la base MCYT-100, et de 41,67% à un seuil de 40 pour la base Philips.

Dans la suite, nous poursuivons notre analyse de notre catégorisation personnalisée d'imitations par rapport à la précédente, en étudiant son impact sur les différents types d'imitations présents dans la base Philips : imitations statiques, dynamiques et professionnelles. Ainsi, nous avons représenté sur la Figure 6.7 des histogrammes illustrant le pourcentage d'imitations de "bonne", "moyenne" et "mauvaise" qualité pour chaque type d'imitations.

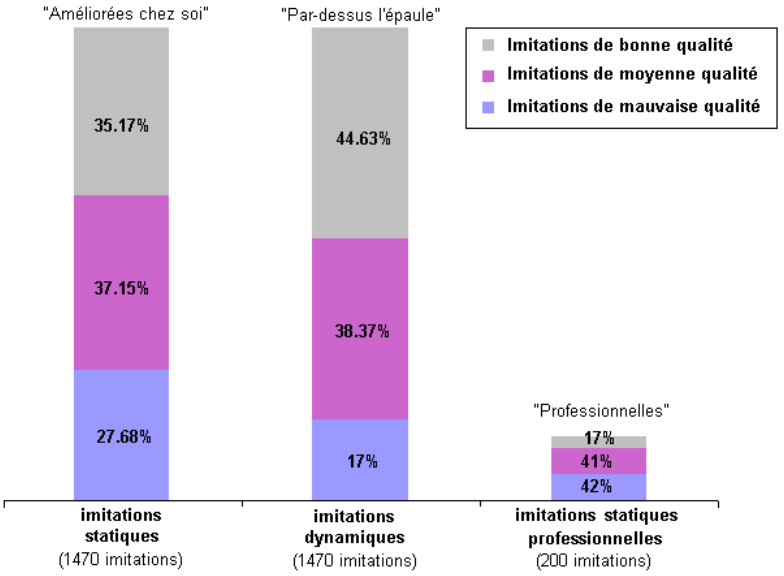

FIGURE 6.7: Pourcentage d'imitations de "bonne", "moyenne" et "mauvaise" qualité pour chaque type d'imitations de la base Philips : statiques, dynamiques et professionnelles.

On retrouve sur la Figure 6.7 les mêmes tendances que sur la Figure 6.4 où l'on avait seulement 2 catégories. En effet, sur la Figure 6.7, on remarque d'abord que le pourcentage le plus élevé des imitations de "bonne" qualité est de nouveau obtenu sur les imitations dynamiques ("par-dessus l'épaule") : 44,63% de ces imitations sont de "bonne" qualité , suivant notre mesure de qualité.

Deuxièmement, on remarque que les imitations "améliorées chez soi" suivent loin derrière en termes de qualité : un tiers seulement de ces imitations statiques sont de "bonne" qualité (35,17%). En outre, on observe que ces imitations contiennent plus d'imitations de"mauvaise" qualité (27,68%) que les imitations dynamiques (17%). En ce qui concerne les imitations professionnelles, on observe qu'avec la catégorisation raffinée, ce type d'imitations contient le plus faible pourcentage d'imitations de "bonne" qualité (17%) et le pourcentage le plus élevé

d'imitations de "mauvaise" qualité (42%), comme déjà observé avec la précédente catégorisation sur la Figure 6.4.

6.3 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle mesure consacrée à la quantification de la qualité des imitations. Cette mesure est basée sur l'extension de notre ancienne mesure d'Entropie Personnelle au contexte des imitations. En fait, comme l'Entropie Personnelle est basée sur les propriétés statistiques locales des signatures authentiques d'une personne estimées par le biais d'un Modèle de Markov Caché, nous avons exploité ces informations statistiques de la personne cible pour mesurer, comme qualité de l'imitation, combien cette dernière va coïncider à la fonction de densité de probabilité de la personne cible. Ainsi, la qualité d'une imitation est quantifiée par la différence existant entre l'Entropie Personnelle associée à la personne et l'entropie de l'imitation, calculée avec le même *MMC* de cette personne.

Afin d'étudier la pertinence de notre mesure de qualité des imitations, nous l'avons comparée aux performances d'un système de vérification basé sur la distance élastique. À cette fin, nous avons d'abord classé les imitations de chaque personne en fonction de leur mesure de qualité en 2 catégories, à savoir imitations de "bonne" et de "mauvaise" qualité. Ensuite, nous avons analysé le comportement de ces catégories en termes de performances dans 3 contextes :

1. Sur la base MCYT-100 qui contient 100 personnes et seulement des imitations statiques.

2. Sur la base Philips qui contient des imitations statiques, dynamiques et professionnelles. Le but de cette expérience est d'analyser le comportement de notre mesure de qualité en fonction de différents types d'imitations.

3. Sur les sous-ensembles BioSecure DS2 et DS3 contenant les mêmes 120 personnes, acquis respectivement sur une plateforme fixe et sur une plateforme mobile. Le but de cette expérience est d'étudier l'impact du protocole et des conditions d'acquisition sur la qualité des imitations.

Dans tous les contextes décrits ci-dessus, nous avons démontré l'efficacité de notre mesure de qualité : le *FAR* est beaucoup plus élevé sur les imitations de "bonne" qualité que sur les imitations de "mauvaise" qualité.

Dans le deuxième contexte, en confrontant notre mesure de qualité aux différents types d'imitations de la base Philips, nous avons abouti aux mêmes résultats que ceux obtenus précédemment dans la littérature par *Dolfing* [23, 24] qui s'est servi d'un classifieur basé sur un *MMC* : dans le contexte en-ligne, les imitations statiques sont de moins bonne qualité que les imitations dynamiques.

Dans le troisième contexte, notre mesure de qualité a montré l'importance des conditions d'acquisition pour la réalisation d'imitations de "bonne" qualité. Notre étude sur les sous-ensembles BioSecure DS2 et DS3 a confirmé qu'avec une interface d'acquisition d'imitations permettant de fournir à un imposteur des informations à la fois sur la forme et la dynamique de la signature cible, on pouvait produire des imitations de "bonne" qualité sur une plateforme mobile. En outre, les imitations de DS3 sont effectivement de meilleure qualité que les imitations statiques de DS2. Ce résultat va dans le même sens que notre précédent résultat sur la base Philips : l'introduction de l'aspect dynamique de la signature cible dans le protocole de production des imitations améliore la qualité des imitations qui en résultent.

Par ailleurs, une catégorisation des imitations personnalisée en 3 ou 4 catégories en fonction des personnes entraîne les mêmes tendances observées avec la catégorisation préliminaire, tout en permettant une meilleure séparation entre les courbes *FAR* des catégories extrêmes contenant respectivement des imitations de "bonne" qualité et de "mauvaise" qualité.

Finalement, notre nouvelle mesure permet d'évaluer la qualité des imitations des bases de signatures en-ligne qui sont à la disposition de la communauté scientifique. Notre étude a montré que les bases MCYT-100, Philips et BioSecure DS2 et DS3 contiennent un pourcentage élevé d'imitations de "bonne" qualité. C'est un résultat important, car ces bases de signatures sont largement utilisées dans la littérature pour l'évaluation des performances des systèmes automatiques de vérification de signature en-ligne.

Chapitre 7

Les différentes applications de la mesure d'Entropie Personnelle

Ce chapitre a pour but de souligner l'intérêt applicatif potentiel de notre mesure de qualité des signatures authentiques basée sur l'Entropie Personnelle, en mettant en évidence sa contribution à l'amélioration de la fiabilité de systèmes de vérification, en particulier en phase opérationnelle, et cela de différentes façons.

Nous commencerons ce chapitre par présenter la campagne d'évaluation de signatures en-ligne BSEC'2009 [25, 52], où les systèmes soumis ont été évalués par catégories de personnes générées avec la mesure d'Entropie Personnelle. Dans la suite de ce chapitre, nous confronterons la mesure d'Entropie Personnelle à des données signature acquises dans différentes conditions (changement de plateforme d'acquisition) [42]. Puis, nous montrerons comment la mesure d'Entropie Personnelle contribue à accroître la fiabilité des systèmes de vérification : d'une part, par la proposition d'un critère à l'enregistrement basé sur la notion de catégories de personnes [43] ; d'autre part, par la proposition d'une méthode permettant la sélection des meilleures signatures de référence et la détection des signatures aberrantes pouvant introduire des erreurs lors de la vérification. Enfin, nous étendrons notre travail par la prise en compte en plus des coordonnées, d'autres fonctions temporelles acquises sur la tablette graphique. Nous étudions la pertinence de ces fonctions temporelles dans le contexte de la variabilité temporelle avec la mesure d'Entropie Personnelle [49].

7.1 Campagne d'évaluation BSEC'2009

Récemment, une nouvelle compétition internationale de vérification de signatures manuscrites en-ligne (BSEC'2009 - BioSecure Signature Evaluation Campaign) a eu lieu en 2009, organisée par notre équipe Intermedia à Telecom sud-Paris [25, 52]. Dans ce qui suit, nous allons présenter brièvement le contexte de l'évaluation BSEC'2009 et ses enjeux, ainsi que les importants résultats qui en découlent. Cette compagne d'évaluation est décrite en détail dans l'Annexe B, où tous les résultats figurent.

7.1.1 Motivation et objectifs

Lors des précédentes compétitions internationales, en l'occurrence la première Compétition Internationale de Vérification de Signatures en-ligne (SVC'2004) [122] et la Campagne d'Evaluation Multimodale de BioSecure (BMEC'2007) [51], les données signatures provenaient d'un seul capteur : une tablette graphique dans le cas de SVC'2004 [122] ; et un PDA dans le cas de BMEC'2007 [51]. D'autre part, les performances ont toujours été évaluées globalement sur toute la base des signatures, et ce indépendamment de toute mesure de qualité. De plus, l'impact de la variabilité temporelle n'a pas été étudié dans les deux compétitions.

La campagne d'évaluation BSEC'2009 [25, 52] comprend plusieurs tâches avec des objectifs distincts, en considérant deux bases de signatures en-ligne acquises dans des conditions bien différentes mais contenant les mêmes personnes. La première base de données est la base BioSecure DS3 [54, 90], dont les signatures ont été acquises sur une plateforme mobile (PDA) ; la deuxième base de données est la base BioSecure DS2 [54, 90], dont les signatures ont été acquises sur une plateforme fixe (tablette graphique). L'objectif de ces deux scénarios d'acquisition est de mesurer l'impact réel des conditions d'acquisition mobiles sur les performances des algorithmes, en utilisant les deux plus grandes bases de signatures en-ligne existantes contenant les mêmes personnes (BioSecure DS2 et DS3 [54, 90]).

Par ailleurs, le problème de la variabilité temporelle des signatures a longuement été étudié dans la littérature, mais à chaque fois sur différentes bases de signatures, avec différents classifieurs et extractions de caractéristiques. Ainsi, BSEC'2009 a pour deuxième objectif l'étude de l'impact de la variabilité temporelle sur les performances des systèmes de vérification et l'évaluation de la pertinence relative au cours du temps des fonctions temporelles capturées par le capteur

[25, 52]. Les bases BioSecure DS2 et DS3 [54, 90] sont en effet bien adaptées à l'étude de cette problématique puisqu'elles ont été collectées en deux sessions espacées dans le temps de plusieurs semaines (se référer à la description de ces deux bases à la Section 2.4.4).

De surcroît, la mesure d'Entropie Personnelle a été validée dans nos travaux de recherche en termes de performances de deux systèmes de vérification développés par notre équipe, l'un basé sur un Modèle de Markov Caché (MMC) et l'autre sur la distance élastique (DTW) [43, 49, 42]. Il nous a paru intéressant, à ce niveau, de présenter à la communauté scientifique le concept de catégories de personnes générées par la mesure d'Entropie Personnelle, et de les valider en termes de performances de plusieurs systèmes de vérification en utilisant la même base de signatures et le même protocole d'évaluation. Ainsi, le troisième et principal objectif de BSEC'2009 [25, 52] est d'évaluer les performances de différents algorithmes en fonction du contenu d'information dans les signatures en utilisant notre mesure d'Entropie Personnelle.

7.1.2 Les participants

Après l'annonce de BSEC'2009 [25, 52], 9 universités de différents pays ont montré leur intérêt à cette campagne d'évaluation. Certaines équipes ont participé aux trois tâches avec un ou plusieurs systèmes, d'autres qu'à la deuxième tâche. Finalement, 14 systèmes ont été soumis. Le Tableau 7.1 indique les équipes qui ont participé à BSEC'2009, ainsi que les tâches auxquelles elles ont participé.

ID	Université	Participants	Tâches
1	Escola Universitaria Politecnica de Mataro, Espagne	J.Roure Alcobé	1-2-3
2		J. Fabregas, M. Faundez-Zany	
3	Institut. de recherche U1, Hongrie	Z.T. Kardkovàcs	1-2-3
4	Univ. de Seikei, Japan	D. Maramatsu	1-2-3
5	Univ. Ain Shams, Egypte	M.I. Khalil, M. Mostafa, H. Abbas	1-2-3
6	Univ. de Valladolid, Espagne	J. M. Pascual-Gaspar, V. Cardeñoso-Payo, C. Vivaracho-Pascual	1-2-3
7	Univ. de Sabanci, Turquie	A. Kholmatov, B. Yanikoglu	1-2-3
8	Univ. Autonoma de Madrid, Espagne	M. Martinez-Diaz, J. Fierrez, J. Ortega-Garcia	1-2-3
9			
10			
11			
12			
13	Univ. de Waseda, Japon	T. Matsumoto	2
14	Univ. de Magdebourg, Allemagne	M. Biermann, T. Scheidat	2
Ref	Telecom SudParis, France	Système de référence	1-2-3

TABLEAU 7.1: Liste des équipes participant à la campagne d'évaluation BSEC'2009.

7.1.3 Bases de développement et de test

Deux ensembles de développement contenant les signatures de 50 personnes appartenant aux deux bases BioSecure DS2 et DS3 [54, 90] ont été distribués aux participants pour développer leurs systèmes avant soumission. Ces deux ensembles, dénommés DS2-50 et DS3-50 contiennent les mêmes 50 personnes.

Après la soumission des systèmes, ces derniers ont été évalués sur deux ensembles de tests dénommés DS2-382 et DS3-382, contenant les signatures de 382 personnes de BioSecure DS2 et DS3 respectivement. A noter, que ces deux ensembles de test sont tenus séquestrés et contiennent les mêmes 382 personnes.

7.1.4 Description des différentes tâches d'évaluation

Suivant les objectifs de BSEC'2009 mentionnés précédemment, l'évaluation des différents systèmes a été effectuée en trois tâches. Nous ne décrivons dans cette partie que les deux tâches qui mettent en pratique la mesure d'Entropie Personnelle, à savoir la $1^{ère}$ et la $3^{ème}$ tâche.

En fait, la Tâche 1 a pour but d'étudier l'impact des conditions d'acquisition sur les performances des systèmes. Nous verrons plus tard dans ce chapitre (Section 7.2) que l'Entropie Personnelle permet de quantifier la dégradation des performances dues au changement de plateforme et confirme ainsi les résultats obtenus dans la Tâche 1. Dans la Tâche 3, les systèmes sont évalués sur les catégories de personnes générées par la mesure d'Entropie Personnelle.

7.1.4.1 Tâche 1 : impact des conditions d'acquisition sur les performances des systèmes

- Les coordonnées sont les seuls paramètres considérés dans cette tâche ;
- Les participants ont optimisé leur système sur la base de développement DS2-50, acquise sur une tablette graphique [54, 90].

Protocole d'évaluation : Les systèmes développés sur DS2-50 ont été testés par l'organisateur sur DS2-382 et DS3-382 pour étudier l'impact des conditions d'acquisition sur les performances des systèmes. Les tests s'effectuent uniquement sur la Session 1.

7.1.4.2 Tâche 3 : impact du contenu d'information dans les signatures sur les performances des systèmes

Protocole d'évaluation : Les systèmes développés sur DS2-50 soumis à la Tâche 1 ont été testés par l'organisateur sur DS2-382 pour chaque catégorie de personnes. Ces catégories ont été générées sur DS2-382 en utilisant une classification hiérarchique ascendante [89] en considérant les valeurs d'Entropie Personnelle des 382 personnes. Les tests ont été effectués uniquement sur la Session 1 de la base DS2-382 en ne considérant que les coordonnées en entrée des systèmes.

7.1.5 Résultats de la Tâche 1

Dans cette tâche, 12 systèmes ont été soumis. Les systèmes ont été développés sur DS2-50, puis testés sur la Session 1 de DS2-382 et DS3-382 (DS2-382 et DS3-382 contiennent les mêmes personnes).

Les Figures 7.1 et 7.2 montrent les courbes de performance obtenues respectivement sur DS2-382 et DS3-382 avec les deux types d'imitations (bonnes et aléatoires). Les taux d'erreur à l'*EER* sont indiqués dans le Tableau 7.2.

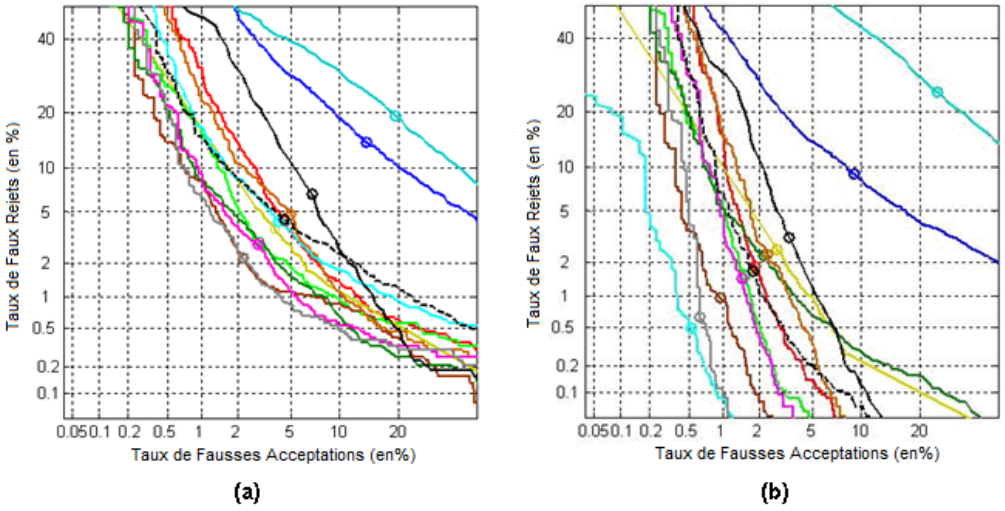

FIGURE 7.1: Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382 en considérant (a) les bonnes imitations et (b) les imitations aléatoires.

Comme on pouvait s'y attendre, les performances sur DS2-382 sont globalement meilleures que sur DS3-382. En considérant les bonnes imitations, les performances sur DS3-382 se dégradent approximativement par un facteur 2 à l'*EER*. En considérant les imitations aléatoires, la dégradation est moins importante.

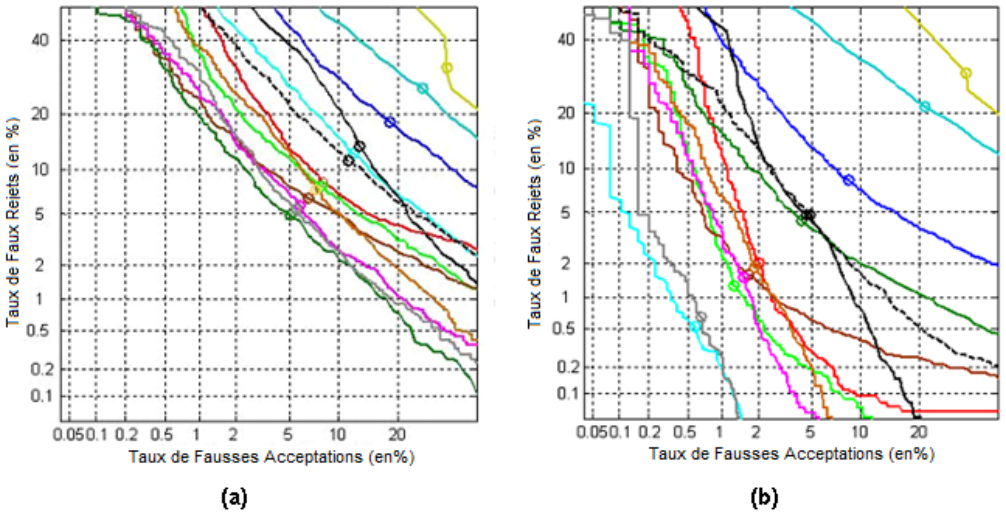

FIGURE 7.2: Courbes de performance des systèmes soumis à la Tâche 1 sur DS3-382 en considérant (a) les bonnes imitations et (b) les imitations aléatoires.

ID	Tâche 1			
	Test sur DS2-382		Test sur DS3-382	
	Bonnes imitations	Imitations aléatoires	Bonnes imitations	Imitations aléatoires
1	4,4%	1,85%	8,18%	2,05%
2	4,91%	2,33%	7,38%	1,86%
3	13,99%	8,98%	18,32%	8,36%
4	2,88%	1,58%	7,87%	1,29%
5	3,82%	2,67%	31,57%	30,64%
6	2,20%	0,97%	6,58%	1,65%
7	2,98%	2,23%	4,99%	4,32%
8	4,18%	0,51%	12,20%	0,55%
9	2,88%	1,47%	5,77%	1,54%
10	19,23%	24,14%	25,85%	21,34%
11	6,71%	3,31%	13,26%	4,7%
12	2,23%	0,63%	5,47%	0,66%
Ref	4,47%	1,74%	11,27%	4,8%

TABLEAU 7.2: Taux d'erreur à l'*EER* des systèmes soumis à la Tâche 1 sur la Session 1 de DS2-382 et DS3-382 avec les deux types d'imitations.

Les mauvaises performances obtenues sur DS3-382 sont dues d'une part à la dégradation de la qualité des signatures acquises sur le PDA, comme nous allons le démontrer plus tard dans la Section 7.2 avec notre mesure d'Entropie Personnelle [42] ; et d'autre part, à la bonne qualité des imitations dans la base BioSecure DS3 comme nous l'avons montré avec notre mesure de qualité des imitations dans la Section 6.2.2.3 du Chapitre 6. En effet, les imitations disponibles dans la base BioSecure DS3 [90, 54] ont été acquises suivant un protocole spécifique : l'imposteur visualise sur l'écran tactile du PDA la séquence d'écriture de la signature cible, puis signe directement sur l'image représentant la trajectoire du stylet (voir la description de la base DS3 dans la Section 2.4.4.2 du Chapitre 2).

7.1.6 Résultats de la Tâche 3

Pour cette évaluation, nous avons considéré les deux catégories extrêmes seulement. La première correspond à la catégorie "Haute Entropie". Elle contient les signatures de 60 personnes. La deuxième correspond à la catégorie "Basse Entropie". Elle contient les signatures de 161 personnes.

Les Figures 7.3 et 7.4 montrent les courbes de performance obtenues sur la Session 1 de DS2-382 pour les deux catégories extrêmes d'Entropie, avec les deux types d'imitations (bonnes et aléatories). Les taux d'erreur à l'*EER* sont indiqués dans le Tableau 7.3.

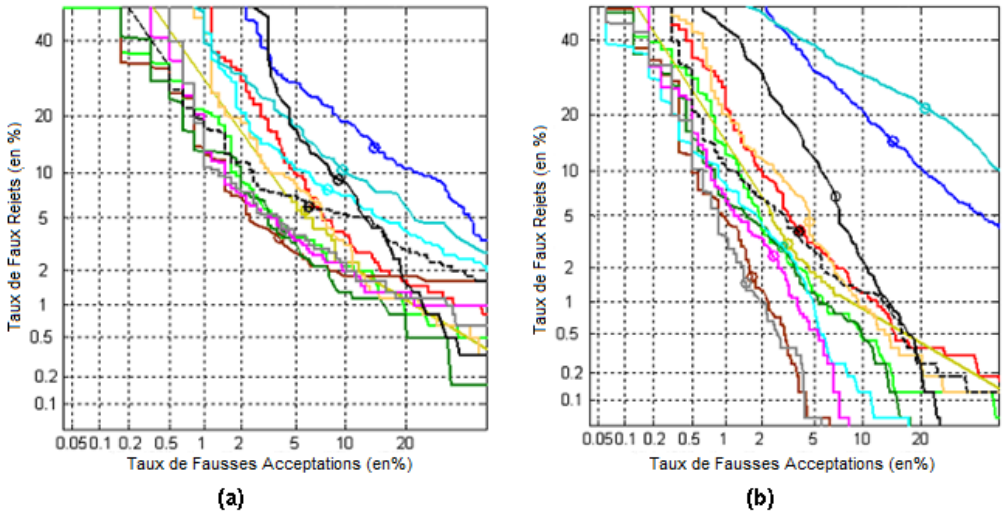

FIGURE 7.3: Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382, en considérant les bonnes imitations, pour les 2 catégories extrêmes d'Entropie : (a) Haute et (b) Basse Entropie.

Hormis les deux systèmes qui ont montré des problèmes lors de l'évaluation BSEC'2009 (*ID* 2 et *ID* 10), pour tous les autres systèmes soumis, les résultats montrent qu'il existe une différence significative des performances entre les deux catégories extrêmes, pour les deux types d'imitations (bonnes et aléatoires) : les meilleures performances sont obtenues sur les personnes appartenant à la catégorie "Basse Entropie", dont les signatures sont les plus longues, les plus complexes et les plus stables. A l'opposé, les performances se dégradent sur les personnes appartenant à la catégorie "Haute Entropie", dont les signatures sont les plus courtes, les moins complexes et les plus variables. Nous avons toutefois noté que certains systèmes sont plus robustes que d'autres à cette dégradation de qualité.

Ces résultats viennent confirmer sur de nombreux systèmes de vérification et de grandes bases de signatures les résultats que nous avons obtenu avec différentes

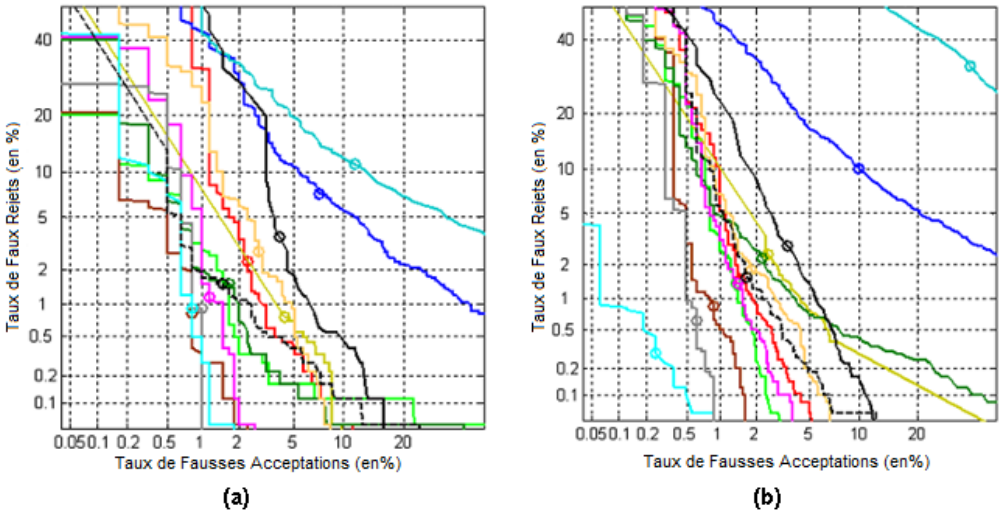

FIGURE 7.4: Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382, en considérant les imitations aléatoires, pour les 2 catégories extrêmes d'Entropie : (a) Haute et (b) Basse Entropie.

ID	DS2-382			
	Haute Entropie		Basse Entropie	
	Bonnes imitations	Imitations aléatoires	Bonnes imitations	Imitations aléatoires
1	6,5%	2,33%	3,94%	1,50%
2	6,58%	2,83%	4,57%	1,80%
3	14,00%	7,22%	14,5%	9,98%
4	4,08%	1,52%	2,92%	1,38%
5	5,67%	2,55%	3,14%	2,47%
6	3,75%	0,83%	1,68%	0,87%
7	4,00%	1,61%	2,89%	2,27%
8	7,83%	0,8%	2,95%	0,27%
9	4,17%	1,19%	2,48%	1,42%
10	9,92%	11,27%	21,18%	32,43%
11	9,00%	3,83%	6,83%	3,14%
12	4,17%	0,91%	1,49%	0,62%
Ref	6,00%	1,52%	3,81%	1,62%

TABLEAU 7.3: Taux d'erreur à l'*EER* sur la Session 1 de DS2-382 pour les deux catégories extrêmes de personnes avec les deux types d'imitations.

approches de vérification au préalable [49, 43, 42]. Ainsi, la capacité de notre mesure d'Entropie Personnelle à catégoriser les utilisateurs apparaît clairement.

7.2 Quantification de la dégradation de qualité des signatures due au changement de plateforme

Comme mentionné dans la Section 1.4.1.3, les conditions de capture peuvent altérer considérablement la qualité du signal et induire ainsi des distorsions dans le tracé de la signature.

En effet, contrairement au cas de l'acquisition d'une signature sur une tablette graphique, cas qui est assez proche de l'utilisation traditionnelle du stylo sur papier assurant ainsi un geste naturel, signer sur une plateforme mobile (PDA) dont la surface d'acquisition est très petite et sur laquelle on ne doit pas poser sa main, entraîne une dégradation de la qualité de signature. De fait, d'après les résultats de la Tâche 1 de l'évaluation BSEC'2009 (Section 7.1.5), nous avons trouvé que les performances des systèmes sur la base DS2 acquise sur une tablette graphique sont meilleures que celles sur la base DS3 acquise sur un PDA.

Il serait alors intéressant de pouvoir affirmer ces résultats par une mesure permettant de caractériser la dégradation de la qualité des signatures quand les personnes changent de plateforme (d'une plateforme fixe à une plateforme mobile). Ainsi, l'objectif fixé dans cette partie est d'analyser l'impact du passage d'une plateforme fixe à une plateforme mobile en termes de notre mesure d'Entropie Personnelle.

Pour cela, nous utilisons les deux ensembles de la base BioSecure DS2-104 et DS3-104 [54, 54], contenant les mêmes 104 personnes, dont les signatures sont acquises sur une plateforme fixe et une plateforme mobile respectivement (voir la description de ces deux bases dans la Section 2.4.4). Puis, nous calculons pour chaque personne de DS2-104 et DS3-104 son Entropie Personnelle en utilisant ses 10 signatures authentiques.

La Figure 7.5 montre les deux histogrammes des valeurs d'Entropie Personnelle sur DS2-104 et DS3-104.

On note sur la Figure 7.5.a que la plupart des personnes (90,28%) de DS2-104 ont une Entropie Personnelle inférieure à 2 bits par seconde, alors que sur la Figure 7.5.b associée à DS3-104, la distribution de l'Entropie Personnelle s'étend sur plus de valeurs. De plus, la valeur maximale d'Entropie Personnelle atteinte

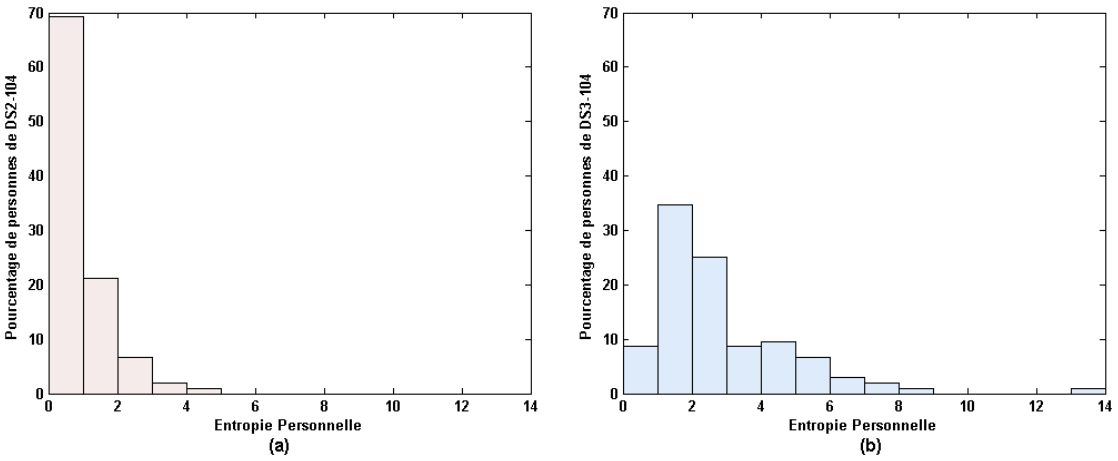

FIGURE 7.5: Distribution des valeurs d'Entropie Personnelle sur les personnes de (a) DS2-104 et (b) DS3-104.

sur DS2 est de 5 bits par seconde, alors que sur DS3 l'Entropie Personnelle atteint une valeur maximale de 14 bits par seconde.

Ainsi, le changement de plateforme de la tablette graphique au PDA a induit une augmentation des valeurs de l'Entropie Personnelle. Cela signifie que les signatures acquises sur une plateforme mobile présentent plus de chaos que celles acquises sur une plateforme fixe. Par conséquent, une dégradation de la qualité des signatures est observée sur la plateforme mobile (PDA).

Pour une analyse plus fine de l'influence du changement de plateforme sur la qualité des signatures, nous étudierons dans la suite l'impact des conditions d'acquisition sur la catégorisation de personnes dans DS2-104 et DS3-104.

Pour cela, nous avons catégorisé séparément, par le biais d'une classification hiérarchique [89], les personnes de DS2-104 et DS3-104 en trois categories suivant leurs valeurs d'Entropie Personnelle.

Le Tableau 7.4 montre la distribution des personnes de DS2-104 et DS3-104 pour chaque catégorie d'Entropie.

Bases de signatures	Catégories d'Entropie Personnelle		
	Haute Entropie	Moyenne Entropie	Basse Entropie
DS2-104	6,73%	25	68,27
DS3-104	16,34%	30,77%	52,88%

TABLEAU 7.4: Pourcentage des personnes de DS2-104 et DS3-104 dans chaque catégorie d'Entropie Personnelle.

On remarque que la distribution des personnes en catégories d'Entropie est différente sur DS2-104 et DS3-104 : le nombre de personnes dans la catégories

"Haute Entropie" a augmenté dans la base DS3-104. Alors que le nombre de personnes dans la catégorie "Basse Entropie" a diminué dans la base DS3-104.

En effet, certaines personnes de DS2-104 ont changé de catégorie d'Entropie quand elles ont signé sur le PDA : 1,92% et 17,30% des personnes de DS2-104 appartenant à la catégorie "Basse Entropie" passent aux catégories "Haute" et "Moyenne Entropie", quand elles signent sur un PDA. Aussi, 8,65% des personnes de DS2-104 appartenant à la catégorie "Moyenne Entropie" passent à la catégorie "Haute Entropie" quand elles signent sur un PDA. On constate aussi qu'une personne de DS2-104 appartenant à la catégorie "Haute Entropie" passe à la catégorie "Moyenne Entropie" quand elle signe sur un PDA.

En conséquence, 14,42% des personnes de DS2-104 sont passées d'une catégorie à une autre de plus grande valeur d'Entropie. Ce résultat signifie que les signatures de DS2, acquises sur une plateforme fixe, deviennent moins complexes et plus variables lorsqu'elles sont acquises sur une plateforme mobile.

La Figure 7.6 illustre la complexité des signatures des 104 personnes lorsqu'elles sont acquises sur une plateforme fixe (DS2-104) en fonction de leur complexité quand elles sont acquises sur une plateforme mobile (DS3-104). Cette mesure de complexité est calculée de la même manière que dans le Chapitre 4 à la Section 4.5.1.

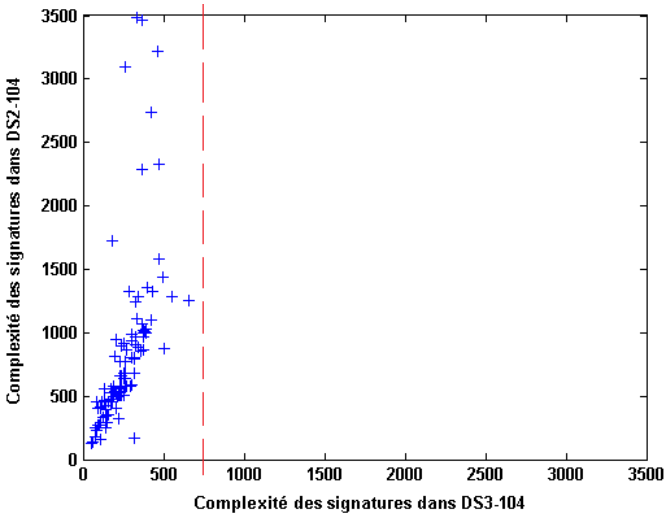

FIGURE 7.6: Complexité des signatures des 104 personnes lorsqu'elles sont acquises sur une plateforme fixe (DS2-104) en fonction de leur complexité quand elles sont acquises sur une plateforme mobile (DS3-104).

On remarque que les valeurs de complexité des signatures des personnes dans DS2-104 diffèrent de celles obtenues avec les mêmes personnes dans DS3-104. En

effet, la complexité des signatures acquises dans DS2-104 sont en moyenne 3,18 fois plus grandes que celles acquises dans DS3-104. Ainsi, signer sur un PDA a tendance a raccourcir les signatures, en les rendant plus variables et moins complexes, comme montré sur la Figure 7.7 par des exemples de signatures de deux personnes qui changent de catégorie d'Entropie Personnelle.

FIGURE 7.7: Signatures de deux personnes qui changent de catégorie sur DS2 et DS3 ; (a) de Moyenne (à gauche) à Haute Entropie (à droite), et (b) de Basse (à gauche) à Moyenne Entropie (à droite).

Finalement, notre mesure d'Entropie Personnelle reflète la dégradation de la qualité des signatures suivant les conditions d'acquisition, puisque l'impact du changement d'une plateforme fixe à une plateforme mobile est mesuré en termes d'augmentation de l'Entropie Personnelle.

7.3 Critère à l'enregistrement

Nous avons montré dans le Chapitre 4 qu'il y a une catégorie de signataires qui sont plus faciles à caractériser que d'autres et sur lesquels les systèmes de vérification font moins d'erreurs, ceux ayant une faible valeur d'Entropie Personnelle. Alternativement, il existe une autre catégorie contenant des signataires problématiques extrêmement difficiles à caractériser et sur lesquels les systèmes de vérification font plus d'erreurs, ceux ayant une grande valeur d'Entropie Personnelle. De plus, les performances des systèmes de vérification dépendent fortement de la qualité des signatures à l'acquisition. Ces constatations doivent être prises en compte lors de l'utilisation de la signature en-ligne dans diverses applications pratiques.

Il serait alors intéressant d'exploiter notre mesure de qualité d'Entropie Personnelle lors de la phase d'acquisition des signatures afin d'adapter la qualité des signatures d'enregistrement au niveau de sécurité demandé par l'application, et cela avant toute étape de vérification [43].

Dans cette optique, nous détaillons dans cette partie un critère à l'enregistrement basé sur la mesure d'Entropie Personnelle pour affiner la procédure d'enregistrement des personnes, et ainsi rendre plus fiable le processus de vérification [43]. Ainsi, en se basant sur notre critère, l'objectif est de pouvoir :

(i) informer l'utilisateur du risque intrinsèque lié à sa propre signature.

(ii) dans le cas d'une signature à risque, laisser la possibilité au signataire de choisir entre poursuivre l'enregistrement sachant le risque intrinsèque à sa signature ou de changer sa signature pour des raisons de sécurité.

(iii) ajuster la qualité de l'enregistrement des signatures au niveau de sécurité requis par l'application.

En fait, le critère à l'enregistrement que nous proposons est basé sur la catégorisation de personnes générées par notre mesure d'Entropie Personnelle au moment de l'enregistrement [43]. Dans ce qui suit, nous allons montrer que sur la base de "Prototypes d'Entropie" générés au préalable sur une base de taille limitée, on peut catégoriser de nouvelles personnes à l'enregistrement, et ainsi mesurer le niveau de sécurité intrinsèque lié à leur signature dès l'étape d'enregistrement.

7.3.1 Extension des catégories de personnes avec l'Entropie Personnelle de DS2-104 à DS2-382

Comme mentionné dans le Chapitre 4, les catégories de personnes sont générées automatiquement avec notre mesure d'Entropie Personnelle par classification hiérarchique [89]. Compte tenu de ce fait, chaque catégorie de personnes est par construction associée à un "Prototype d'Entropie" hérité de la classification hiérarchique. Chaque "Prototype d'Entropie" correspond en effet à la moyenne des valeurs d'Entropie Personnelle associées aux personnes appartenant à la catégorie considérée [43].

Notre objectif dans cette partie est d'étudier la possibilité de catégoriser de nouvelles personnes en exploitant les "Prototypes d'Entropie" générés au préalable sur un ensemble de signatures de taille limitée. Nous montrons que cela est possible à certaines conditions : les signatures des nouvelles personnes doivent être acquises avec le même capteur d'acquisition, à la même résolution et suivant le même protocole d'acquisition que celles de l'ensemble de taille limitée.

Cette étude est effectuée en générant trois "Prototypes d'Entropie" sur la base BioSecure DS2-104, contenant 104 personnes. Puis, ces trois prototypes sont utilisés pour catégoriser les nouvelles personnes appartenant à un autre ensemble de la base BioSecure de plus grande taille : DS2-382 contenant les signatures de 382 personnes. L'ensemble DS2-382 est le même que celui utilisé pour le test dans BSEC'2009 [25, 52].

Ainsi, chaque personne appartenant à l'ensemble DS2-382 est classée suivant sa valeur d'Entropie Personnelle par la méthode d'extension des prototypes [43] en adoptant le protocole suivant :

1. Récupérer les trois "Prototypes d'Entropie" (un par catégorie d'Entropie Personnelle) calculés au préalable sur l'ensemble DS2-104. Ces prototypes correspondent à la moyenne des valeurs d'Entropie Personnelle dans chacune des 3 catégories d'Entropie ;

2. Calculer l'Entropie Personnelle de chaque personne de la base DS2-382 en se servant de leurs 10 signatures authentiques ;

3. Suivant la méthode du plus proche voisin [8], associer à chaque personne de DS2-382 la catégorie d'Entropie dont le "Prototype" est le plus proche à la valeur d'Entropie associée à la personne considérée.

Le Tableau 7.5 montre la distribution des personnes de DS2-382 en chaque catégorie d'Entropie ainsi générées.

Statistiques	Catégories de personnes de DS2-382		
	Haute Entropie	Moyenne Entropie	Basse Entropie
Pourcentage de personnes	15,71%	42,15%	42,14%
Prototype d'Entropie	3,19	1,47	0,49

TABLEAU 7.5: Distribution des personnes de DS2-382 dans chaque catégorie d'Entropie générée suivant les "Prototypes d'Entropie" calculés au préalable sur DS2-104.

On observe que même dans le cas d'une très grande base de signatures, les catégories "Moyenne" et "Basse Entropie" contiennent beaucoup plus de personnes que la catégorie "Haute Entropie". En effet, cette dernière contient à peu près 2,7 fois moins de personnes que les deux autres catégories.

Afin d'analyser la pertinence de ce protocole de catégorisation, nous étudierons dans la partie suivante les performances d'un système de vérification sur les catégories de personnes de DS2-382 ainsi obtenues. L'étude des performances est effectuée

avec un classifieur basé sur les Modèles de Markov Cachés (MMCs), le même que celui utilisé dans le Chapitre 4, en considérant les deux types d'imitations.

7.3.2 Evaluation des performances par catégorie de personnes de DS2-382

La Figure 7.8 et le Tableau 7.6 montrent les performances du classifieur MMC avec les deux types d'imitations (bonnes et aléatoires) sur chaque catégorie de personnes de la base DS2-382, obtenues après avoir calculé les "Prototypes d'Entropie" sur DS2-104. Pour un but comparatif, nous avons aussi tracé les performances globales obtenues sur toute la base DS2-382, toutes catégories confondues.

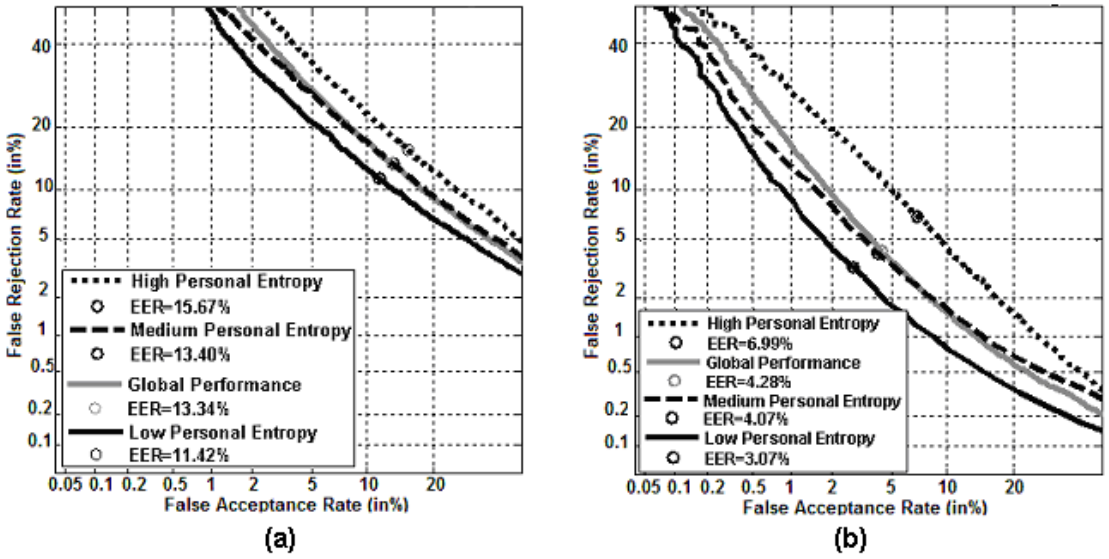

FIGURE 7.8: Courbes de performance sur toute la base DS2-382 et sur chaque catégorie de personnes de DS2-382 avec le classifieur MMC, en considérant (a) les bonnes imitations et (b) les imitations aléatoires.

DS2-382	Bonnes imitations		Imitations aléatoires	
	EER (%)	IC (95%)	EER (%)	IC (95%)
Haute Entropie	15,67	$\pm 0,025$	6,99	$\pm 0,015$
Moyenne Entropie	13,40	$\pm 0,012$	4,07	$\pm 0,006$
Basse Entropie	11,42	$\pm 0,014$	3,07	$\pm 0,005$
Performance globale	13,34	$\pm 0,003$	4,28	$\pm 0,003$

TABLEAU 7.6: Taux d'erreur à l'*EER* et intervalles de confiance pour chaque catégorie de personnes de DS2-382 avec le classifieur MMC en considérant les deux types d'imitations (bonnes et aléatoires).

Comme observé dans le Chapitre 4 sur la base DS2-104, sur laquelle les "Prototypes d'Entropie" ont été calculés, on remarque aussi sur la base DS2-382 qu'il

existe une différence dans les performances du classifieur MMC entre les catégories, pour les deux types d'imitations : les plus mauvaises performances sont obtenues sur les personnes appartenant à la catégorie "Haute Entropie". A l'opposé, les performances s'améliorent significativement sur les personnes appartenant à la catégorie "Basse Entropie".

D'après le Tableau 7.6, on note qu'à l'*EER*, les performances sont améliorées d'un facteur de 2 pour les bonnes imitations et de 1,4 pour les imitations aléatoires quand on passe de la catégorie "Haute Entropie" à la plus basse. On remarque aussi que la catégorie "Moyenne Entropie" donne des performances intermédiaires entre celles obtenues avec les deux catégories extrêmes.

Par ailleurs, comme on pouvait s'y attendre, on constate que les performances globales sur toute la base DS2-382 sont mauvaises en comparaison aux performances de la catégorie "Basse Entropie", mais se rapprochent beaucoup des performances de la catégorie "Moyenne Entropie".

Jusqu'à présent, les résultats obtenus suivant la procédure d'extension des prototypes sont cohérents avec ceux du Chapitre 4. Ainsi, ces résultats attestent de la validité de notre protocole de catégorisation basé sur l'extension des prototypes.

Cependant, il est à signaler que pour une bonne extension, la base sur laquelle les "Prototypes d'Entropie" sont calculés doit être assez représentative de la population et doit bien sûr contenir des signatures des 3 catégories.

D'autre part, la procédure d'extension des prototypes est importante lorsqu'il s'agit de catégoriser les personnes dans des bases de grande taille parce que la classification non supervisée est très sensible à la taille de l'échantillon à catégoriser.

Dans ce qui suit, nous décrivons la nouvelle procédure d'enregistrement proposée basée sur l'extension des prototypes pour catégoriser de nouvelles personnes.

7.3.3 Proposition d'un critère à l'enregistrement

D'après nos précédents résultats, les signatures les moins fiables en termes de performance présentent de très grandes valeurs d'Entropie Personnelle [43, 42, 49]. De plus, nous avons montré que les "Prototypes d'Entropie" générés au préalable sur une base de taille limitée (DS2-104) peuvent être utilisés pour effectuer la catégorisation de nouvelles personnes acquises dans les mêmes conditions que la base de taille limitée [43].

Par conséquent, nous proposons d'exploiter ces "Prototypes d'Entropie", qui sont totalement indépendants du système de vérification, pour introduire un critère à l'enregistrement reposant sur l'identification au préalable des signatures les moins fiables en termes de performances par le biais de la mesure d'Entropie Personnelle.

La procédure d'enregistrement de personnes que nous proposons comporte les étapes suivantes :

1. Dix signatures authentiques sont demandées au signataire à l'enregistrement ;

3. L'Entropie Personnelle associée à ce signataire est calculée sur ses 10 signatures authentiques ;

4. Les trois "Prototypes d'Entropie" calculés au préalable sur une base de taille limitée (DS2-104) sont récupérés ;

5. Le signataire est affecté à l'une des trois catégories d'Entropie, celle dont le "Prototype" est le plus proche de la valeur d'Entropie Personnelle de ce signataire par la méthode du plus proche voisin [8].

Par la suite, trois cas de figure peuvent être rencontrés :

1. Si le signataire est classé comme appartenant à la catégorie "Haute Entropie", il doit être informé du risque intrinsèque lié à sa signature. En effet, cette catégorie de personnes est peu fiable par rapport aux autres catégories d'Entropie. Nous proposons alors au signataire l'alternative de changer sa signature pour des raisons de sécurité ou de poursuivre l'enregistrement sachant le risque intrinsèque lié à sa signature.

2. Si le signataire est classé comme appartenant à la catégorie "Basse Entropie", il est directement enregistré, car cette catégorie de personnes est la plus fiable.

3. Si le signataire est classé comme appartenant à la catégorie "Moyenne Entropie", nous lui recommandons de réaliser une signature plus complexe et moins variable, mais il peut toujours garder sa signature.

Ainsi, la procédure d'enregistrement que nous avons proposée ci-dessus et en [43] repose sur la mesure d'Entropie Personnelle. Pour compléter cette procédure, il serait intéressant d'envisager, dans le cas où des imitations sont disponibles, l'utilisation de la mesure d'Entropie Relative Personnelle introduite dans le Chapitre 5 qui tient aussi compte de la vulnérabilité d'une signature aux attaques.

7.4 Sélection des signatures de référence

Dans le cas de la vérification d'identité, le système compare la signature acquise sur le capteur au moment du test à celles servant de référence pour l'identité proclamée, ainsi le système infirme ou confirme l'identité proclamée. Il s'avère alors très important que les signatures servant de référence, comme leur nom l'indique, soient les plus représentatives possible de la personne afin de limiter les erreurs du système lors de la vérification.

De plus, comme la signature manuscrite d'une personne varie d'un instant à l'autre, chaque signataire a une variabilité intra-classe qui lui est propre. En effet, les signatures de référence de deux personnes ne varient pas de la même manière d'un instant à l'autre : certaines personnes présentent plus de variabilité que d'autres. Plus cette variabilité est forte, plus le système de vérification va rejeter des signatures authentiques ou accepter des imitations.

Il est ainsi important de sélectionner pour chaque personne les signatures de référence les plus représentatives de sa variabilité intra-classe intrinsèque et qui assurent le plus faible taux d'erreur lors de la vérification. A ces fins, il faut tenir compte de la qualité intrinsèque de chaque signature authentique de la personne enregistrée et ne prendre comme signatures de référence que celles qui sont de bonne qualité.

Ainsi, nous proposons dans cette partie de sélectionner les meilleures signatures de référence en se basant sur notre critère d'Entropie Personnelle, qui comme nous l'avons montré, caractérise un signataire en termes de complexité et de variabilité de ses signatures [43].

Nous montrerons dans ce qui suit qu'en nous servant de la mesure d'Entropie Personnelle, calculée suivant un protocole spécifique, nous pouvons sélectionner la meilleure base de références contenant les 5 signatures authentiques les plus représentatives de la personne, ainsi que la pire base de références. Nous verrons que cela a un impact important sur les performances du système de vérification.

7.4.1 Protocole proposé pour la sélection des signatures de référence

Dans les bases de signatures, plusieurs exemplaires authentiques N sont disponibles par personne. Pour sélectionner les meilleures signatures de référence pour

chaque personne, nous calculons la mesure d'Entropie en suivant un protocole de type "leave-one-out" [73]. Ce protocole consiste en une succession de découpages avec *N-1* signatures en apprentissage (toutes sauf une) pour tester sur la signature restante.

Ainsi, on estime les densités de probabilité locales de la signature d'une personne avec un MMC sur ses signatures authentiques moins une. Puis, on calcule l'Entropie Personnelle H_p associée à la personne en moyennant les valeurs d'Entropie sur les *N-1* signatures authentiques utilisées pour l'apprentissage du MMC. Avec ce même MMC, on calcule aussi *H(sig)* l'Entropie de la signature *"sig"* restante qui ne faisait pas partie de la base d'apprentissage des densités de probabilité locales.

Par la suite, on calcule la dissimilarité Q_{sig} existant entre l'Entropie de la signature $H(sig)$ et l'Entropie Personnelle calculée sur les *N-1* signatures H_p :

$$Q_{sig} = |H(sig) - H_p| \tag{7.1}$$

Ce processus est répété autant de fois qu'il y a de signatures authentiques par personne. Autrement dit, ce processus est répété N fois pour les N signatures authentiques. A chaque signature authentique sera donc associée une valeur de dissimilarité Q_{sig}.

A la fin, on ordonne par ordre croissant les N signatures authentiques de la personne suivant leur valeur de dissimilarité Q_{sig}. Les signatures authentiques ayant les plus faibles valeurs de dissimilarité d'Entropie sont les signatures jugées appropriées pour être des signatures de référence pour notre système de vérification. Nous analysons à la suite l'impact de cette procédure de sélection des références sur la performance d'un système de vérification.

7.4.2 Etude des performances du système de vérification

Cette étude de sélection des signatures de référence est effectuée sur la base MCYT-100 [91]. Les signatures sont représentées uniquement par les coordonnées *(x,y)*.

Comme MCYT-100 contient 25 signatures authentiques par personne, l'estimation des densités de probabilité locales par le MMC est effectuée sur 24 signatures authentiques. L'Entropie Personnelle est calculée sur ces 24 signatures

et l'Entropie de la signature restante est aussi calculée. Ce processus est répété pour toutes les 25 signatures authentiques (c.à.d. 25 fois). Enfin, pour chaque personne de MCYT-100, à chacune des 25 signatures authentiques leur est associée une valeur de dissimilarité d'Entropie Q_{sig} calculée suivant l'équation 7.1.

Pour l'étude des performances, nous considérons deux bases de référence de qualité extrême : la meilleure et la pire. Comme dans un système de vérification, le nombre de signatures de référence usuel est de 5, la meilleure base de références est celle contenant les 5 signatures authentiques parmi les 25 ayant les plus faibles valeurs de dissimilarité d'Entropie ; la pire base de références est celle contenant les 5 signatures authentiques parmi les 25 ayant les plus fortes valeurs de dissimilarité d'Entropie.

7.4.2.1 Description du classifieur et protocole d'évaluation

Afin d'étudier la pertinence de notre méthode de sélection des 5 signatures de référence, on évalue la performance d'un classifieur basé sur la distance élastique (DTW) selon le protocole d'évaluation suivant : on considère pour chaque personne de MCYT-100 la meilleure base de références en premier temps, puis la pire base de références en second temps. Le test est effectué sur les 15 signatures authentiques restantes (n'appartenant pas aux deux bases de références sélectionnées, à savoir la meilleure et la pire), 15 bonnes imitations et 15 imitations aléatoires. Le score DTW utilisé est calculé comme suit :

$$Score_{DTW} = \frac{1}{N}\sum_{i=1}^{N} DTW(test, reference_i) \tag{7.2}$$

où N est le nombre de signatures de référence, et $Score_{DTW}$ est la distance moyenne calculée entre la signature de test et les 5 signatures de référence.

Dans un but comparatif, on considère aussi pour chaque personne deux bases de références dont les 5 signatures authentiques ont été tirées aléatoirement, comme souvent pratiqué pour l'évaluation des systèmes de vérification dans la littérature.

7.4.2.2 Résultats

La Figure 7.9 montre les courbes de performance du classifieur DTW sur la base MCYT-100 avec les deux types d'imitations (bonnes et aléatoires), en

considérant dans un cas les meilleures signatures de référence par personne, dans l'autre les pires, puis les deux cas de signatures de référence tirées aléatoirement. Les différents taux d'erreur à l'*EER* sont donnés dans le Tableau 7.7.

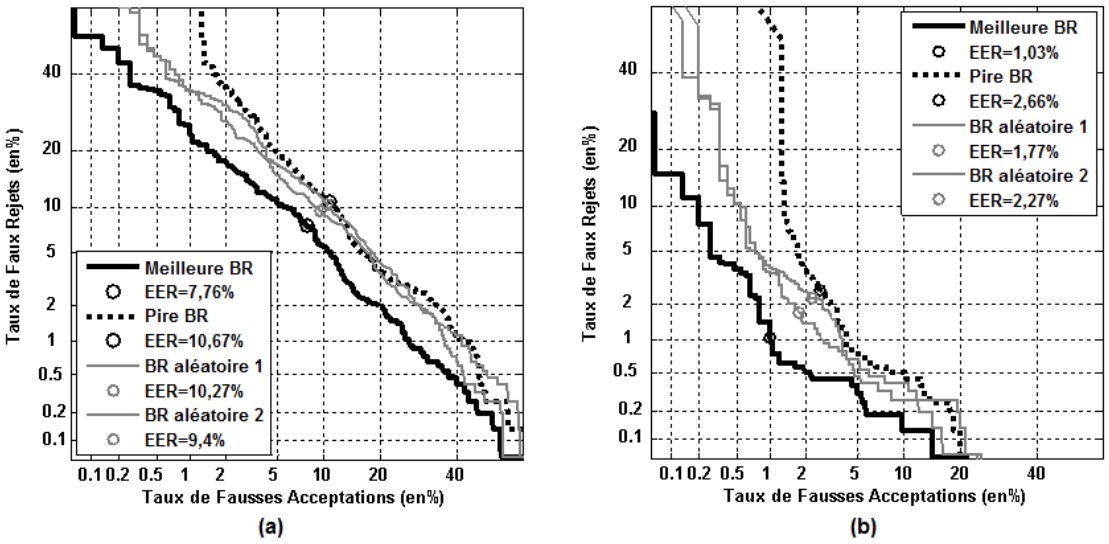

FIGURE 7.9: Courbes de performance du classifieur DTW sur les différentes bases de référence en considérant (a) les bonnes imitations et (b) les imitations aléatoires.

Bases de références	Bonnes imitations	Imitations aléatoires
La meilleure	**7,76%**	**1,03%**
Aléatoire 1	10,27%	1,77%
Aléatoire 2	9,4%	2,27%
La pire	**10,67%**	**2,66%**

TABLEAU 7.7: Taux d'erreur à l'*EER* du classifieur DTW sur chaque base de références en considérant les deux types d'imitations (bonnes et aléatoires).

On observe que les meilleures performances sont en effet obtenues avec la meilleure base de références, au sens de notre critère de dissimilarité d'Entropie entre chaque référence et l'ensemble des signatures authentiques, celle contenant les 5 signatures dont les valeurs de dissimilarité d'Entropie sont les plus faibles. A l'opposé, les pires performances sont obtenues avec la pire base de références, celle contenant les 5 signatures dont les valeurs de dissimilarité d'Entropie sont les plus fortes.

En effet, d'après le Tableau 7.7, l'amélioration relative des performances avec la meilleure base de références en comparaison à la pire est de 27,27% à l'*EER* avec les bonnes imitations et de 61,27% à l'*EER* avec les imitations aléatoires. Cette amélioration est observée à l'*EER* mais aussi sur tous les points de fonctionnement de la courbe. Elle est d'autant plus accentuée aux faibles taux de *FA*. En fait,

notre méthode de sélection identifie les signatures de référence qui sont les plus représentatives de la personne, c'est-à-dire celles qui présentent la plus grande similarité en termes d'Entropie (et donc de complexité et de stabilité) à toutes les autres signatures authentiques réalisées par cette personne. Ce principe de sélection des meilleures signatures de référence aura donc pour effet une diminution certaine et significative du taux de faux rejets, et cela est nettement observé dans la zone de haute sécurité correspondant aux faibles valeurs de FA (lorsque le seuil de décision est très faible).

Par ailleurs, on remarque que les signatures de référence choisies aléatoirement donnent des performances intermédiaires entre celles obtenues avec les deux bases de références extrêmes.

Dans cette partie, nous avons présenté une méthode de sélection des signatures de référence basée sur la mesure d'Entropie Personnelle. Nous avons montré l'efficacité de cette méthode en la confrontant aux performances d'un système de vérification. Cependant, il faut aller encore plus loin dans l'analyse des résultats en comparant cette méthode à celles utilisées dans la littérature pour la sélection des signatures de référence [76, 77, 85, 2, 85].

Il serait aussi intéressant d'envisager l'utilisation de la mesure d'Entropie Relative Personnelle, dans le cas où des imitations sont disponibles, pour la sélection des signatures de référence les plus représentatives de la personne et les plus robustes aux attaques.

7.5 Détection des signatures aberrantes

Dans le même esprit que la précédente étude relative à la sélection des signatures de référence, il arrive parfois à l'enregistrement qu'une personne rate sa propre signature, ou bien qu'à cause d'un problème de capture, la signature acquise présente des aberrations. Sans mesure de qualité, on ne peut pas détecter cette signature considérée comme une "anomalie" et qu'il est toutefois important de détecter parce qu'elle peut induire des erreurs importantes lors du processus de vérification.

Ainsi, nous proposons dans cette partie une méthode pour identifier les signatures aberrantes dans une base de données, la base MCYT-100 [91] dans notre cas, en nous servant de notre mesure d'Entropie Personnelle calculée suivant le même protocole ("leave-one-out" [73]) décrit dans la Section 7.4.1.

Néanmoins, la première difficulté réside dans la définition de ce qu'est une "signature aberrante", car certaines personnes présentent une forte variabilité intra-classe et dans ce cas une signature peut être très différente des autres signatures authentiques sans qu'elle doive pour autant être considérée comme une signature ratée.

7.5.1 Protocole proposé pour la détection des signatures aberrantes

Cette étude est effectuée sur la base MCYT-100 [91]. Les signatures sont représentées uniquement par leurs coordonnées *(x,y)*.

Selon le même protocole décrit dans la Section 7.4.1, on estime les densités de probabilité locales avec un Modèle de Markov Caché (MMC) sur toutes les signatures authentiques de la personne moins une. Ensuite, on calcule l'Entropie des *N-1* signatures considérées pour l'apprentissage du MMC, ainsi que l'Entropie de la signature de la personne restante $H(sig)$ avec le même MMC. Puis, on calcule la moyenne H_p et l'écart-type H_{sigma} de l'Entropie sur toutes les *N-1* signatures.

Par la suite, afin de trouver une frontière entre la variabilité intrinsèque de la personne et une signature "ratée", on normalise l'Entropie de la signature restante en calculant la différence entre l'Entropie de cette signature $H(sig)$ et la moyenne H_p , rapportée à l'écart-type H_{sigma} :

$$Q'_{sig} = \frac{|H(sig) - H_p|}{H_{sigma}} \tag{7.3}$$

Ce processus est répété autant de fois qu'il y a de signatures authentiques par personne. A chaque signature authentique sera ainsi associée une valeur de dissimilarité d'Entropie normalisée Q'_{sig}.

L'idée est que les signatures authentiques présentant de fortes valeurs de dissimilarité d'Entropie normalisée Q'_{sig} seront considérées comme des signatures aberrantes.

7.5.2 Résultats

La Figure 7.10 montre les valeurs de dissimilarité d'Entropie normalisée Q'_{sig} pour toutes les 2500 signatures authentiques disponibles dans la base MCYT-100.

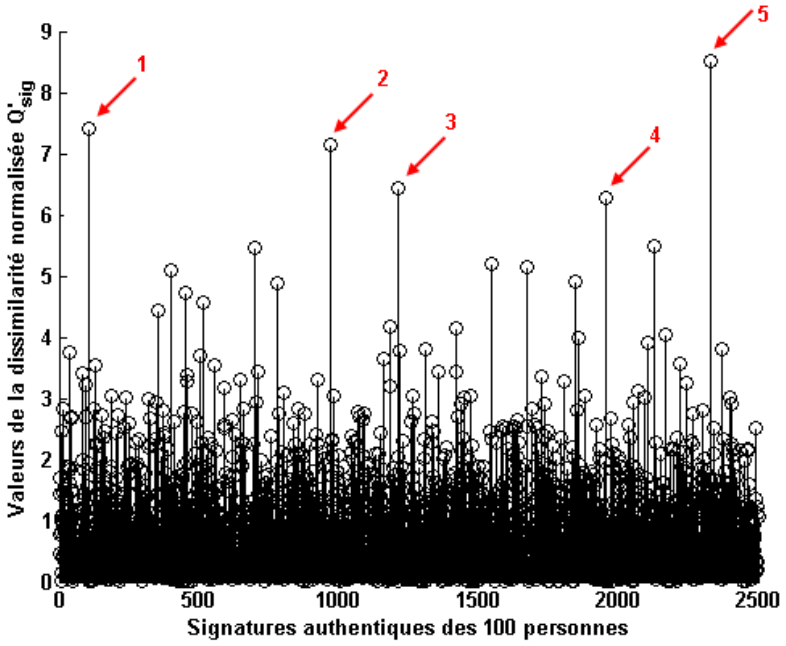

FIGURE 7.10: Valeurs de la dissimilarité normalisée Q'_{sig} pour toutes les 2500 signatures authentiques disponibles dans la base MCYT-100.

On remarque que la plupart des signatures authentiques de MCYT-100 présentent des valeurs de dissimilarité d'Entropie normalisée inférieures à 4, alors que certaines signatures authentiques présentent de plus fortes valeurs.

A ce niveau d'étude, des questions se posent alors : parmi ces signatures présentant de fortes valeurs, lesquelles sont aberrantes ? Quel est le seuil à partir duquel une signature doit être considérée comme aberrante ?

Pour s'affranchir du seuil, nous avons procédé à une classification non supervisée hiérarchique ascendante [89] appliquée sur les valeurs de dissimilarité d'Entropie normalisée associées aux 2500 signatures authentiques de MCYT-100.

La Figure 7.11 présente le dendrogramme [89], appelé aussi arbre hiérarchique, issu de la classification hiérarchique appliquée sur les valeurs de dissimilarité d'Entropie normalisée des 2500 signatures authentiques de MCYT-100.

Nous avons présenté aussi à droite de la Figure 7.11 les niveaux de nœuds de l'arbre (en noir), qui correspondent précisément à un intervalle entre deux des itérations (agrégations) du processus de classification hiérarchique ascendante [89].

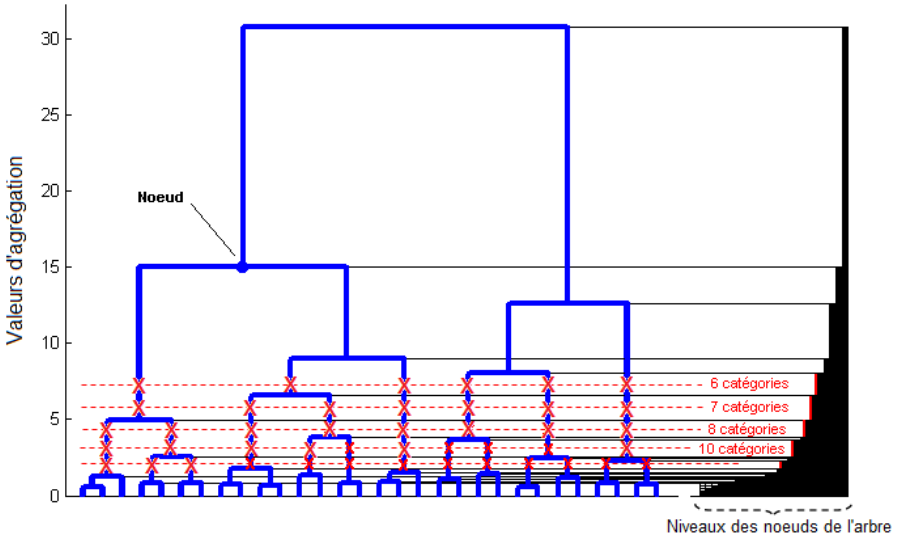

FIGURE 7.11: Dendrogramme de la classification hiérarchique ascendante.

L'idée est de faire des catégories de signatures, et ne considérer que la catégorie extrême dont les valeurs de dissimilarité d'Entropie Q'_{sig} sont les plus fortes.

Une solution pratique pour avoir une idée sur le nombre de catégories de signatures est d'observer la distance entre les agrégations successives dans le diagramme des niveaux de nœuds : en partant des faibles valeurs d'agrégation, on cherche les premières grandes distances d'agrégation, c.à.d. les premiers écarts en hauteur entre deux nœuds successifs ; puis en chaque palier correspondant à ces écarts de distance on coupe l'arbre hiérarchique par une droite horizontale, comme montré par les lignes horizontales discontinues rouges sur la Figure 7.11. Toute coupure d'une droite horizontale avec l'arbre fournit alors un nombre de partitions possibles. Ainsi, sur la Figure 7.11 on trouve que le nombre de classes possibles peut être 6, 7, 8, 10 ou 13.

En appliquant notre classification hiérarchique avec les différentes possibilités du nombre de classes (6, 7, 8, 10 ou 13), on trouve dans tous les cas que la catégorie extrême, c.à.d. celle dont les valeurs de dissimilarité d'Entropie sont les plus grandes, contient les mêmes 5 exemples de signature. Ces signatures appartiennent d'ailleurs à 5 personnes différentes et sont numérotées et identifiées sur la Figure 7.10 par une flèche rouge.

Ainsi, avec une classification non supervisée, on a pu détecter les signatures aberrantes sans pour autant se préoccuper à chercher une valeur de seuil pour les identifier.

La Figure 7.12 montre des exemples de signatures authentiques d'une personne dont une est une signature aberrante détectée avec notre méthode. A noter que la signature de cette personne a déjà été publiée dans [32, 4].

FIGURE 7.12: Exemple de signature aberrante (anomalie) détectée par notre mesure d'Entropie Personnelle.

Pour un meilleur aperçu concernant la validité de notre méthode de détection des signatures aberrantes, nous avons calculé pour chaque signature authentique d'une personne de la base MCYT-100 la distance élastique (DTW) moyenne entre cette signature et toutes les autres signatures de la personne. Les signatures aberrantes montreront de fortes valeurs de cette distance.

La Figure 7.13 illustre les distances DTW moyennes obtenues pour les 25 signatures de 3 personnes, dont l'une de leur signature a été jugée aberrante suivant notre méthode d'Entropie. Les signatures aberrantes sont indiquées par la flèche rouge et numérotées suivant le même ordre sur la Figure 7.12.

On constate que les signatures considérées aberrantes avec notre méthode d'Entropie montrent de fortes distances DTW. Cela est nettement observé pour les signatures aberrantes n°1 (Figure 7.13.a) et n°4 (Figure 7.13.b). Sur la Figure 7.13.c, la signature aberrante n°3 montre aussi une forte valeur de la distance DTW, néanmoins on remarque que la DTW n'arrive pas à séparer les signatures aberrantes des signatures très variables. Ce qui confirme la capacité de notre mesure d'Entropie à détecter et isoler les signatures aberrantes par rapport à une autre méthode basée sur la distance élastique (DTW).

La recherche des signatures aberrantes a pour intérêt pratique de fiabiliser les performances de différents systèmes de vérification opérationnels. Cependant, pour une application efficace de notre méthode de détection des signatures aberrantes dans des systèmes opérationnels, reste à tester une procédure de généralisation permettant de détecter naturellement toute nouvelle signature aberrante à l'enregistrement.

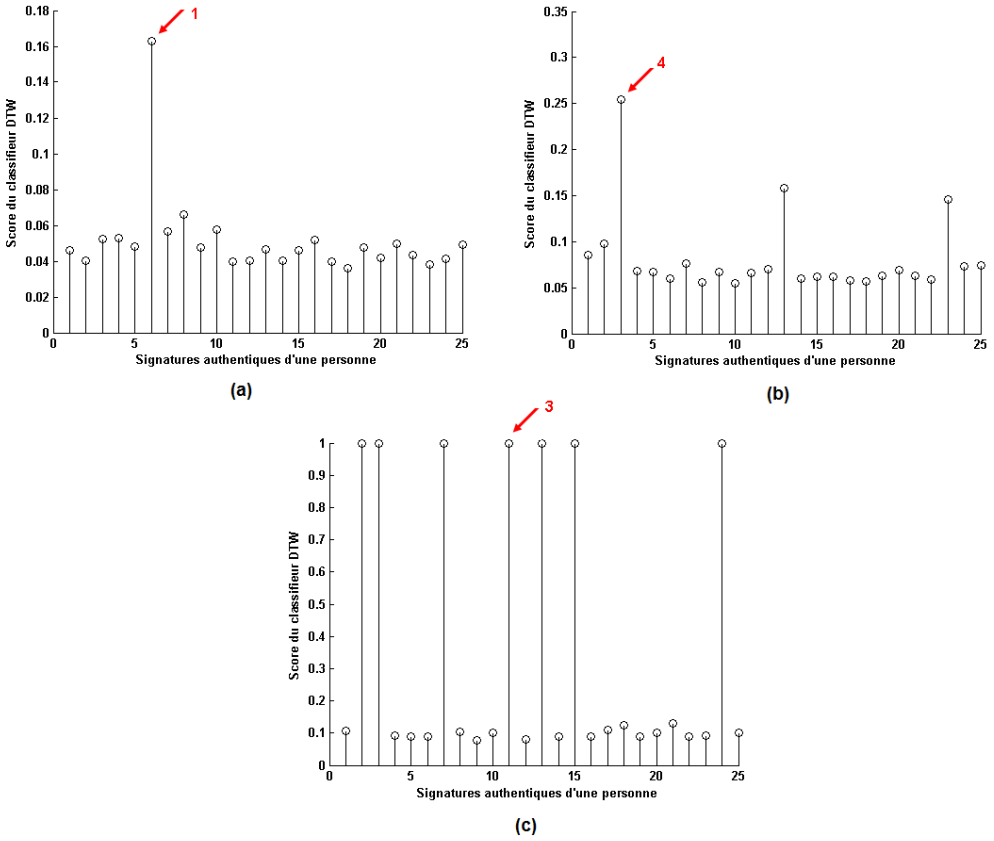

FIGURE 7.13: Signatures aberrantes confrontées au score DTW.

7.6 Etude de la robustesse des fonctions temporelles à la variabilité temporelle

La vérification de la signature en-ligne repose sur une paramétrisation de la signature par différentes fonctions temporelles discrètes acquises sur un capteur à intervalles de temps réguliers : les coordonnées *(x,y)*, la pression *(p)* et les angles d'inclinaison *(Az,Alt)* du stylo (dans le cas d'une tablette graphique). Comme la signature est une modalité comportementale qui présente une forte variabilité au cours du temps, il est alors crucial de choisir la paramétrisation de signature la plus robuste à la variabilité temporelle lorsque le système de vérification est évalué dans ce contexte.

En dépit des différents travaux de la littérature concernant la "meilleure" paramétrisation des signatures [34, 46, 88, 122], la plupart d'entre eux n'ont pas considéré l'aspect variabilité temporelle. Ainsi, l'impact de la variabilité temporelle sur ces fonctions n'a jamais été étudié auparavant dans la littérature. D'autre part, ces travaux montrent des résultats contradictoires : en utilisant un classifieur basé sur la DTW, certains affirment que les angles d'inclinaison donnent les meilleures

performances en comparaison à une représentation de la signature par les coordonnées seulement ou par la pression seulement [46] ; alors que d'autres affirment que les angles d'inclinaison sont des paramètres très instables [34]. Un travail plus récent sur le sujet affirme que les performances du système peuvent être améliorées lorsque les 5 fonctions temporelles sont combinées, mais l'efficacité de cette combinaison dépend du classifieur utilisé [88]. Par ailleurs, les résultats de la compétition SVC'2004 [122] montrent que l'utilisation seule des coordonnées donne de meilleurs résultats que la combinaison des 5 fonctions temporelles *(x,y,p,Az,Alt)*. Toutefois, les travaux de la littérature semblent être d'accord avec le fait que la pression est un paramètre discriminant entre les personnes.

Dans cette partie, notre étude porte sur l'influence relative des fonctions temporelles acquises sur une tablette graphique sur les performances des systèmes de vérification [50], notamment dans un contexte de variabilité temporelle des signatures (les signatures sont acquises sur différentes sessions espacées dans le temps).

Comme dans la plupart des travaux de la littérature, l'efficacité des paramètres dépend du classifieur [88], nous proposons pour notre étude deux critères totalement indépendants de l'étape classification [50]. Le premier repose sur notre mesure d'Entropie Personnelle présentée dans le Chapitre 4, quantifiant à la fois la complexité et la variabilité des signatures ; le deuxième repose sur une mesure de distance qui quantifie la variabilité intra-classe de la personne.

Nous confrontons aussi les résultats obtenus avec ces deux critères aux performances de deux systèmes de vérification. Cette phase permet de constater l'impact de la variabilité temporelle sur les taux d'erreur des systèmes, et ce pour différentes combinaisons de fonctions temporelles.

Pour cette étude, nous distinguons deux types de variabilité temporelle : l'une dite à "court-terme" (il y a quelques minutes d'intervalle entre deux sessions d'acquisition de signatures) et l'autre dite à "long-terme" (il y a quelques semaines d'intervalle entre deux sessions d'acquisition). A ces fins, deux bases de signatures en-ligne sont utilisées : la base MCYT-100 [91] acquise en 5 sous-sessions espacées de quelques minutes ; et la base Biomet [39] acquise en 2 sessions espacées de 5 mois (voir la description des bases dans la Section 2.4 du Chapitre 2). Ce travail requiert alors la définition de protocoles adéquats pour étudier la pertinence des différents paramètres à plusieurs niveaux de variabilité temporelle.

Dans la suite, nous décrirons d'abord les deux critères utilisés pour analyser la robustesse à la variabilité temporelle des différentes combinaisons de fonctions temporelles, à savoir l'Entropie Personnelle et la mesure de variabilité intra-classe. Puis, nous présenterons les différents protocoles et les deux classifieurs utilisés pour l'évaluation des performances en présence de différents contextes de variabilité temporelle. Par la suite, nous comparerons les deux critères avec les performances des systèmes de vérification.

7.6.1 Comparaison des différentes fonctions temporelles par les deux critères

Notre étude [50] est réalisée au moyen des deux critères présentés dans la section suivante. Et ce, pour chaque combinaison de fonctions temporelles à la suite :

- Coordonnées *(x,y)* seulement ;
- Coordonnées et pression *(x,y,p)* ;
- Coordonnées et angles d'inclinaison *(x,y,Az,Alt)* ;
- Les 5 fonctions temporelles combinées *(x,y,p,Az,Alt)*.

7.6.1.1 Premier critère : Entropie Personnelle

Nous proposons d'utiliser la mesure d'Entropie Personnelle pour comparer les différentes fonctions temporelles indépendamment de tout classifieur, en présence ou pas de variabilité temporelle. Dans le contexte d'absence de variabilité temporelle, l'Entropie Personnelle est calculée sur des signatures authentiques appartenant à une seule session ; dans le cas de présence de variabilité temporelle, elle est calculée sur des signatures authentiques appartenant à deux sessions. Et ce, pour chacune des 4 combinaisons citées auparavant [50].

Dans ce cadre, nous exploitons le critère d'Entropie Personnelle comme suit : avec certaines combinaisons de fonctions temporelles, la valeur de l'Entropie Personnelle change en présence de variabilité temporelle, ce qui signifie que ces combinaisons ont tendance à augmenter l'instabilité de la signature au cours du temps. En conséquence, de telles combinaisons de fonctions temporelles ne sont pas robustes à la variabilité temporelle.

7.6.1.2 Deuxième critère : variabilité intra-classe

Pour la mesure de la variabilité intra-classe d'une personne, on calcule la distance élastique [96] entre tous les couples possibles des signatures authentiques de cette personne (on obtiendra par exemple 45 couples lorsque l'on considère 10 signatures authentiques par personne). Puis on moyenne les distances obtenues pour avoir une mesure de variabilité intra-classe associée à la personne.

Comme pour le critère d'Entropie, la variabilité intra-classe est mesurée en considérant des signatures authentiques appartenant à une seule session (absence de variabilité temporelle) ou à deux sessions (présence de variabilité temporelle). Et ce, pour chacune des 4 combinaisons citées auparavant.

Les combinaisons de fonctions temporelles présentant une stabilité de la valeur de variabilité intra-classe, calculée en présence et en absence de variabilité temporelle, seront considérées robustes à la variabilité temporelle.

7.6.2 Protocoles d'expérimentation

Dans cette partie, nous présentons les 3 protocoles utilisés pour notre étude. Pour chaque protocole, nous détaillons deux configurations : une pour l'évaluation des critères et l'autre pour l'évaluation des performances des classifieurs (décrits en 7.6.2.1).

Pour l'étude des performances, nous utilisons deux classifieurs : l'un basé sur la distance élastique (DTW) [96], l'autre sur les Modèles de Markov Cachés (MMCs) [96]. Pour le classifieur DTW, le score de dissimilarité est la distance moyenne entre la signature de test et les 5 signatures de référence. Dans le cas du système basé sur un MMC, le score de similarité utilisé résulte de la fusion par une simple moyenne du score de vraisemblance et du score de segmentation, comme décrit en détail dans [110, 109].

Protocole 1 : absence de variabilité temporelle

Les expériences de cette étude sont effectuées uniquement sur la Session 2 de la base Biomet. Les critères d'Entropie Personnelle et de variabilité intra-classe sont calculés en considérant pour chaque personne 10 signatures authentiques appartenant à la Session 2 de la base Biomet.

Pour l'évaluation des performances, 100 tirages aléatoires sont réalisés sur les signatures authentiques et les imitations de la base Biomet : chaque tirage contient 5 signatures de référence de la Session 2. Pour le test, on considère les 5 signatures authentiques restantes de la Session 2 et les 6 bonnes imitations de la Session 2.

Protocole 2 : présence de variabilité temporelle à "long-terme"

Les expériences de cette étude sont effectuées sur les deux sessions de la base Biomet. Les critères d'Entropie Personnelle et de variabilité intra-classe sont calculés en considérant 15 signatures authentiques par personne (10 de la Session 2 et 5 de la Session 1).

Pour l'évaluation des performances, 100 tirages aléatoires sont réalisés sur les signatures authentiques et les imitations : chaque tirage contient 5 signatures de référence de la Session 2. Pour le test, on considère les 5 signatures authentiques appartenant à la Session 1 et les 6 bonnes imitations de la Session 1.

Pour étudier l'impact de la variabilité temporelle, on comparera les résultats obtenus suivant le Protocole 2 avec ceux obtenus suivant le Protocole 1.

Protocole 3 : présence de variabilité temporelle à "court-terme"

Les expériences de cette étude sont effectuées sur les 5 sous-sessions de la base MCYT-100 [91]. Dans cette étude, le critère de variabilité intra-classe n'est pas utilisé. Le critère d'Entropie Personnelle est calculé d'abord sur les 10 signatures authentiques appartenant aux 2 sous-sessions A et B ensemble, car pour une bonne estimation de l'Entropie, il faut 10 signatures au minimum comme souligné dans le Chapitre 4. Puis, l'Entropie est calculée sur un plus grand ensemble contenant 25 signatures appartenant aux 5 sous-sessions (de A à E). Ainsi, on comparera les résultats sur deux ensembles de signatures espacés au moins de 30 minutes.

Pour l'évaluation des performances, on calcule le taux d'erreur *EER* en considérant pour chaque personne les 5 signatures authentiques de la 1ère sous-session (dénommée A) comme signatures de référence. Pour le test, on considère les 5 signatures authentiques et les 5 bonnes imitations de chacune des 4 sous-sessions (B, C, D, E) séparément. Ceci a pour but de comparer les performances sur des ensembles de tests espacés de quelques minutes.

7.6.3 Analyse des résultats

7.6.3.1 Protocole 1 vs. Protocole 2

Dans cette partie, nous allons étudier la pertinence des 4 combinaisons de fonctions temporelles dans le cadre de variabilité temporelle à "long-terme" présente dans la base Biomet [39], en se basant sur les critères d'Entropie Personnelle et variabilité intra-classe ainsi que sur les performances des deux classifieurs DTW et MMC.

Critère d'Entropie Personnelle

La Figure 7.14 montre pour chacune des 4 combinaisons de fonctions temporelles les valeurs d'Entropie Personnelle pour toutes les personnes de la base Biomet calculées suivant les Protocoles 1 et 2.

A noter que pour chaque combinaison de fonctions temporelles résulte un MMC différent. Ainsi, les valeurs d'Entropie Personnelle de sont pas comparables entre les différentes combinaisons de fonctions temporelles. Le but n'est pas de comparer les fonction temporelles entre elles, mais de comparer les configurations variées de protocoles avec une même combinaison de fonctions temporelles.

On remarque que pour les coordonnées (Figure 7.14.a), les valeurs d'Entropie calculées suivant le Protocole 1 (absence de variabilité temporelle) et suivant le Protocole 2 (présence de variabilité temporelle à "long-terme") se chevauchent pour la plupart des personnes. Ce résultat signifie que la combinaison *(x,y)* est robuste à la variabilité temporelle à "long-terme".

En revanche, les combinaisons contenant les angles d'inclinaison (Figures 7.14.c et 7.14.d) montrent pour la plupart des personnes une forte séparation entre les deux valeurs d'Entropie calculées suivant les Protocoles 1 et 2. Ainsi, les angles d'inclinaison ne sont pas robustes à la variabilité temporelle à "long-terme".

Enfin, on remarque que la combinaison *(x,y,p)* contenant les coordonnées et la pression (Figure 7.14.b) est moins résistante à la variabilité temporelle que celle contenant uniquement les coordonnées *(x,y)*, mais reste plus résistante à la variabilité temporelle que celles contenant les angles d'inclinaison *(x,y,Az,Alt)* et *(x,y,p,Az,Alt)*. En conséquence, suivant le critère d'Entropie Personnelle, *(x,y)* est la combinaison la plus robuste à la variabilité temporelle à "long-terme".

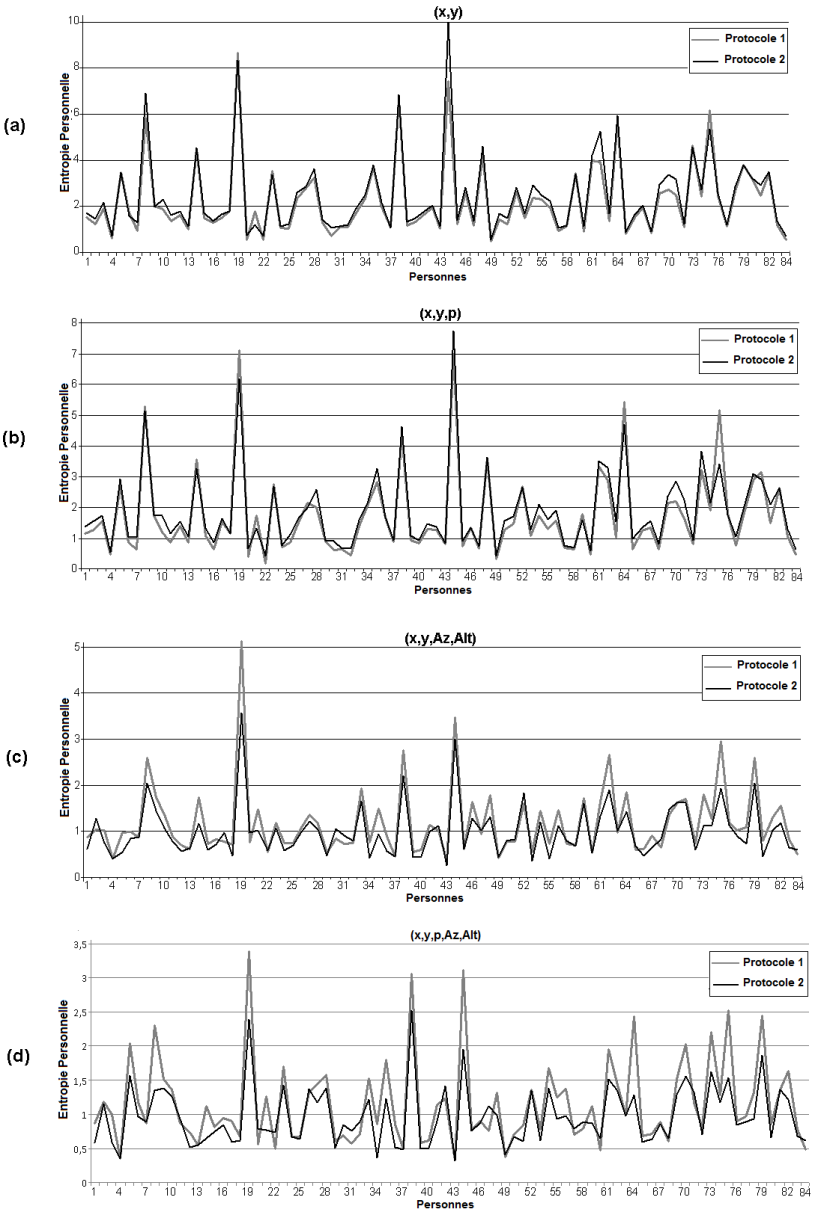

FIGURE 7.14: Valeurs d'Entropie des personnes de Biomet calculées suivant les Protocoles 1 et 2 pour les combinaisons : (a) (x,y), (b) (x,y,p), (c) (x,y,Az,Alt), et (d) (x,y,p,Az,Alt).

Critère de variabilité intra-classe

La Figure 7.15 montre pour chacune des 4 combinaisons les distributions de la variabilité intra-classe calculée suivant les Protocoles 1 et 2, pour toutes les personnes de la base Biomet.

D'emblée, on remarque que les valeurs de la variabilité intra-classe les moins élevées sont celles correspondant aux fonctions temporelles uniquement spatiales (Figure 7.15.a). On remarque aussi une augmentation progressive des valeurs de

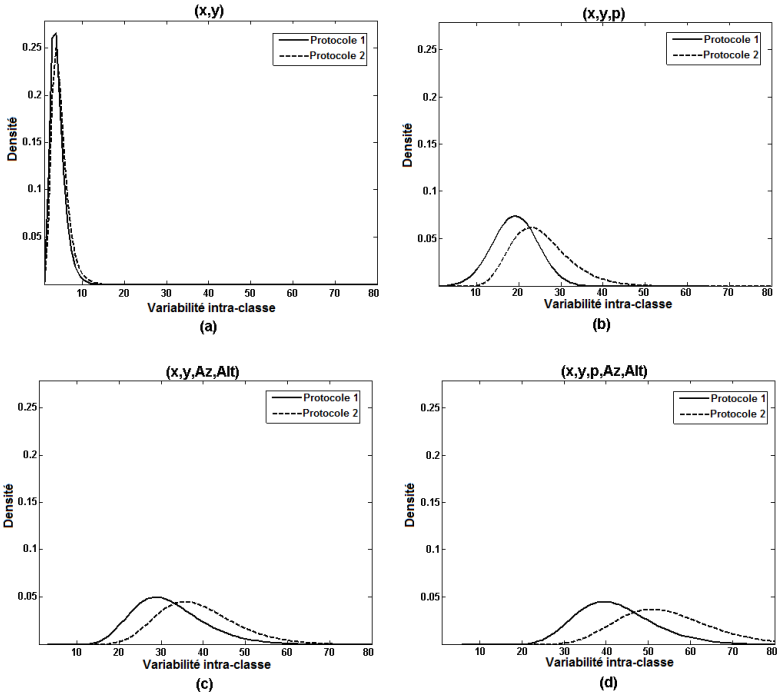

FIGURE 7.15: Distributions de la variabilité intra-classe calculées suivant les Protocoles 1 et 2 sur la base Biomet pour les combinaisons : (a) (x,y), (b) (x,y,p), (c) (x,y,Az,Alt), et (d) (x,y,p,Az,Alt).

variabilité intra-classe lorsque la pression (Figure 7.15.b) puis les angles d'inclinaison (Figures 7.15.c et 7.15.d) sont ajoutés séparément aux coordonnées, au sein d'une même session. De plus, les distributions de variabilité se décalent vers des valeurs plus élevées lorsque les angles d'inclinaison sont présents dans la combinaison considérée.

En outre, on constate que pour *(x,y)* la distribution de la variabilité intra-classe calculée suivant les Protocoles 1 et 2 reste inchangée (voir la Figure 7.15.a), c'est-à-dire sur la même plage de valeurs et avec un même aspect de la distribution sur une ou deux sessions séparées de cinq mois. Ceci vient confirmer les résultats observés avec le critère d'Entropie Personnelle, à savoir que les fonctions temporelles spatiales sont les plus robustes à la variabilité temporelle.

Un deuxième résultat intéressant est le suivant : lorsque la pression est présente dans la combinaison de fonctions temporelles utilisée, la distribution change d'aspect par rapport au résultat obtenu sur une seule session (Protocole 1). En effet, la distribution suivant le Protocole 2 est davantage étalée sur la plage des valeurs des mesures de variabilité intra-classe et sa moyenne augmente. Ceci signifie que la pression accroît notoirement la variabilité intra-personnelle lorsque l'on considère une échelle de temps de plusieurs mois. Ceci vient confirmer les

résultats observés avec le critère d'Entropie Personnelle, à savoir que la pression n'est pas un paramètre robuste à variabilité temporelle à "long-terme".

Pour les combinaisons contenant les angles d'inclinaison, on observe que les deux distributions suivant les Protocole 1 et 2 sont décalées (Figures 7.15.c, 7.15.d). Ainsi, les angles d'inclinaison ne sont pas robustes à la variabilité temporelle à "long-terme".

En conséquence, avec le critère de variabilité intra-classe, on trouve aussi que les coordonnées sont les fonctions temporelles les plus robustes à la variabilité temporelle à "long-terme".

Evaluation des performances

La Figure 7.16 montre les courbes de performance des classifieurs DTW et MMC pour chacune des 4 combinaisons, suivant les Protocoles 1 et 2. Les Tableaux 7.8 et 7.9 indiquent les taux d'erreur à l'*EER* ainsi que les intervalles de confiance.

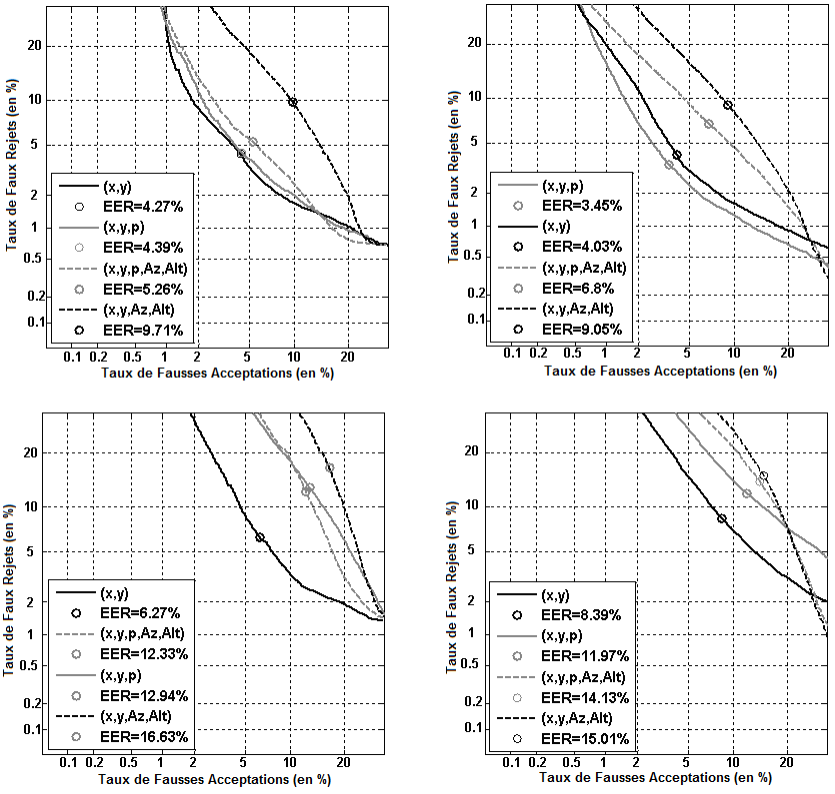

FIGURE 7.16: Courbes de performance sur la base Biomet avec les classifieurs DTW (à gauche) et MMC (à droite) pour chaque combinaison de paramètres, suivant le Protocole 1 (en haut) et le Protocole 2 (en bas).

Combinaisons	Protocole 1			
	DTW		MMC	
	EER	IC (95%)	EER	IC (95%)
(x,y)	4,27%	±0,082	4,03%	±0,74
(x,y,p)	4,39%	±0,081	3,45%	±1,04
(x,y,Az,Alt)	9,71%	±0,125	9,05%	±0,49
(x,y,p,Az,Alt)	5,26%	±0,113	6,8%	±0,12

TABLEAU 7.8: Taux d'erreur à l'*EER* et intervalles de confiance à 95% pour chaque combinaison de paramètres sur la base Biomet avec les classifieurs DTW et MMC suivant le Protocole 1.

Combinaisons	Protocole 2			
	DTW		MMC	
	EER	IC (95%)	EER	IC (95%)
(x,y)	6,27%	±0,11	8,39%	±0,10
(x,y,p)	12,94%	±0,14	11,97%	±0,12
(x,y,Az,Alt)	16,63%	±0,13	15,01%	±0,14
(x,y,p,Az,Alt)	12,33%	±0,13	14,13%	±0,12

TABLEAU 7.9: Taux d'erreur à l'*EER* et intervalles de confiance à 95% pour chaque combinaison de paramètres sur la base Biomet avec les classifieurs DTW et MMC suivant le Protocole 2.

On remarque tout d'abord que la variabilité temporelle introduit une dégradation dans les performances quelle que soit la combinaison de paramètres considérée. On remarque aussi qu'en présence ou pas de variabilité temporelle à "long-terme", la combinaison *(x,y,Az,Alt)* donne les moins bonnes performances avec les deux classifieurs. Ainsi, les angles d'inclinaison ne sont pas recommandés en cas d'absence ou de présence de variabilité à "long-terme".

On constate aussi qu'en l'absence de variabilité temporelle, *(x,y)* donne les meilleures performances avec la DTW, suivie de près par *(x,y,p)*, tandis qu'avec le MMC, *(x,y,p)* est celle qui donne les meilleures performances. Ainsi, le comportement de *(x,y,p)* dépend du classifieur utilisé, comme signalé dans [88].

Cependant, en présence de variabilité temporelle à "long-terme", *(x,y,p)* est nettement moins bonne que *(x,y)* et cela pour les deux classifieurs, notamment avec un taux d'erreur environ 2 fois plus grand que celui engendré par les coordonnées avec le classifieur DTW. En effet, les coordonnées donnent les meilleurs résultats dans ce contexte (voir Tableaux 7.8 et 7.9). Néanmoins, la pression reste le meilleur paramètre à utiliser par rapport aux angles d'inclinaison.

En conséquence, l'étude des performances confirme les résultats obtenus avec les critères d'Entropie et de Variabilité intra-classe : l'aspect spatial de la signature est le plus robuste à la variabilité temporelle à "long-terme".

7.6.3.2 Protocole 3

Dans cette partie, nous étudions la pertinence des fonctions temporelles dans le cadre de variabilité à "court-terme" présente dans la base MCYT-100 [91], avec le critère d'Entropie Personnelle et les performances des deux classifieurs.

Critère d'Entropie Personnelle

La Figure 7.17 montre pour les 4 combinaisons de fonctions temporelles les valeurs d'Entropie Personnelle calculées suivant le Protocole 3 pour toutes les personnes de MCYT-100.

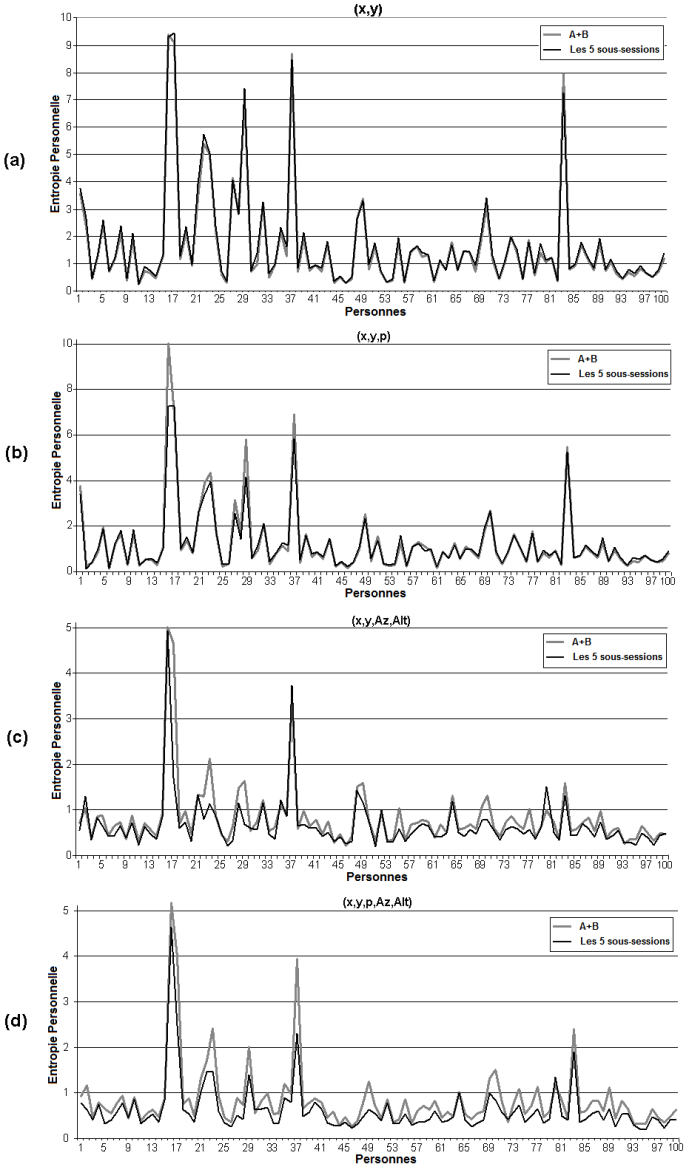

FIGURE 7.17: Valeurs d'Entropie des personnes de MCYT-100 calculées suivant le Protocole 3 avec : (a) (x,y), (b) (x,y,p), (c) (x,y,Az,Alt), et (d) (x,y,p,Az,Alt).

On remarque sur la Figure 7.17.a et la Figure 7.17.b que la valeur d'Entropie Personnelle calculée sur les 10 signatures des deux sous-sessions *(A+B)* ou sur les 25 signatures des cinq sous-sessions (de A à E) reste stable sur toutes les personnes pour la combinaison *(x,y)* et pour presque toutes les personnes pour la combinaison *(x,y,p)*. On constate ainsi que ces combinaisons sont robustes à la variabilité temporelle à "court-terme".

A contrario, on remarque sur les Figures 7.17.c et 7.17.d que lorsque l'on considère les angles d'inclinaison, les valeurs d'Entropie Personnelle changent entre l'ensemble contenant les 2 sous-sessions *(A+B)* et l'ensemble contenant les 5 sous-sessions. Cela montre que les paramètres d'angles d'inclinaison ne sont pas robustes dans le contexte de variabilité à "court-terme", et donc dans aucun des cas considérés par les 3 protocoles.

Critère d'évaluation des performances

Le Tableau 7.10 indique le taux d'erreur *EER* des deux classifieurs DTW et MMC pour les 4 combinaisons de fonctions temporelles sur les ensembles de test B, C, D et E.

Combinaisons	Protocole 3 : DTW				Protocole 3 : MMC			
	A-B	A-C	A-D	A-E	A-B	A-C	A-D	A-E
(x,y)	7,56%	7,6 %	7,84%	10%	11,24%	11,76%	13,92%	14,62%
(x,y,p)	5,2%	5,94%	9,84%	8%	7,86%	8,38%	9,44%	10,88%
(x,y,Az,Alt)	7%	7,18%	7,4%	9,42%	13,22%	18,10%	15,32%	15,88%
(x,y,p,Az,Alt)	5%	5,46%	6,58%	7,6%	10,06%	15,94%	12,6%	12,92%

TABLEAU 7.10: Taux d'erreur à l'*EER* pour chaque combinaison de paramètres sur MCYT-100 avec les classifieurs DTW et MMC suivant le Protocole 3.

En premier lieu, on constate avec la DTW qu'une dégradation progressive des performances apparaît d'autant plus que la session de test est éloignée dans le temps de la session des signatures de référence, et cela pour les 4 combinaisons de fonctions temporelles.

En revanche, avec le MMC, les résultats varient beaucoup lorsque les angles d'inclinaison sont pris en compte, tandis que la tendance reste la même que celle observée avec la DTW pour les combinaisons *(x,y)* et *(x,y,p)*. Ceci confirme la difficulté à modéliser les angles d'inclinaison pour une même personne. D'une manière générale, on remarque que la combinaison *(x,y,Az,Alt)* dégrade les performances des systèmes considérablement par rapport à *(x,y,p)*.

Par ailleurs, les résultats du Tableau 7.10 confirment qu'en général *(x,y,p)* améliore les performances dans le cadre de variabilité à "court-terme" par rapport aux seules fonctions coordonnées. En effet, dans le cadre à "court-terme", on constate que les fonctions coordonnées donnent de moins bons résultats que les combinaisons contenant la pression *(x,y,p)* et *(x,y,p,Az,Alt)* mais, comme montré dans l'expérience en présence de variabilité temporelle à "long-terme" sur Biomet dans la Section 7.6.3.1, les coordonnées sont les plus robustes à la variabilité à "long-terme".

En conclusion, nous trouvons que contrairement à certains travaux de la littérature [46], les angles d'inclinaison sont de mauvais paramètres en présence ou pas de variabilité temporelle. Nous avons en effet montré par les deux critères d'Entropie et de variabilité intra-classe que ces paramètres sont instables dans les contextes de variabilité "court-terme" et "long-terme". De plus, ces fonctions temporelles n'ont pas montré dans l'évaluation des performances une capacité de discrimination des imitées.

Aussi, nous avons observé avec notre mesure d'Entropie et les performances de deux systèmes de vérification que la pression est un bon paramètre dans le contexte de variabilité à "court-terme", comme observé dans la littérature [46, 34, 88]. Toutefois, dans ces travaux, la variabilité temporelle n'a pas été considérée. Nous avons prouvé par plusieurs méthodes (Entropie Personnelle, variabilité intra-classe et performances de deux systèmes) que la pression devient très instable et dégrade les performances dans le contexte de variabilité à "long-terme".

Finalement, bien que les fonctions temporelles spatiales donnent dans certains cas de moins bonnes performances sur une seule session par rapport à *(x,y,p)*, elles restent de loin les plus robustes à la variabilité temporelle à "long-terme".

Ainsi, pour un système de vérification de signatures robuste à la variabilité temporelle, il est préférable de garder les fonctions spatiales et d'en extraire des paramètres dynamiques tels que la vitesse, l'accélération etc., et de ne pas utiliser la pression et les angles d'inclinaison.

7.7 Conclusion

Dans ce chapitre, seule la mesure d'Entropie Personnelle a été considérée.

Nous avons montré que la mesure d'Entropie Personnelle ne sert pas uniquement à classer les personnes suivant la qualité de leur signature, mais à bien d'autres applications toutes aussi importantes contribuant ainsi à l'amélioration des performances des systèmes de vérification.

Chapitre 8

Conclusions et perspectives

Dans cette thèse nous avons proposé de nouvelles mesures de qualité pour les signatures manuscrites en-ligne. Après une présentation générale du contexte de cette thèse dans le Chapitre 1 et des problématiques liées à l'utilisation de la signature en-ligne, nous avons brièvement donné la ligne directrice de ce travail axée sur 3 contributions originales majeures. La première contribution est la proposition d'une mesure permettant de quantifier la qualité des signatures authentiques d'une personne, en termes de complexité et de stabilité [42, 49, 43]. La deuxième contribution consiste en la proposition d'une autre mesure permettant de quantifier la vulnérabilité des signatures aux attaques. La troisième est la proposition d'une mesure permettant de quantifier la qualité d'échantillons d'imitation.

Dans le Chapitre 2, nous avons présenté les notions générales d'un système de vérification d'identité par la signature en-ligne, les outils généraux d'évaluation des performances ainsi que les approches de classification les plus utilisées dans la littérature.

Ensuite, dans le Chapitre 3, nous avons présenté les travaux de la littérature relatifs aux mesures de qualité appliquées à la signature manuscrite. Plusieurs critères de qualité "naturels" ont été utilisés : complexité, variabilité et lisibilité. Ceux-ci sont calculés de manière plus ou moins similaire dans les différents travaux et en général un lien est fait avec les performances des systèmes de vérification. Néanmoins, nous avons noté le manque de travaux axés sur les imitations.

Par la suite, nous avons développé chacune des trois contributions principales de cette thèse. La première contribution, présentée dans le Chapitre 4, est une nouvelle mesure de qualité appelée "Entropie Personnelle", calculée sur un ensemble

de signatures authentiques d'une personne [42, 49, 43]. Cette mesure est calculée en estimant les densités de probabilité localement, sur des portions de la signature, en nous servant d'un Modèle de Markov Caché. L'originalité de cette mesure réside dans le fait qu'elle englobe les critères utilisés dans la littérature, à savoir : la complexité, la stabilité et la lisibilité. Cette mesure permet de générer par classification non supervisée des catégories de personnes à la fois en termes de stabilité de la signature et de complexité du tracé. En la confrontant aux performances des approches de vérification de référence dans la littérature, les expériences ont montré que les performances varient de façon significative (d'un facteur 2 au minimum) entre les personnes de la catégorie de plus "Haute Entropie" (signatures instables et peu complexes) et ceux de la catégorie de plus "Basse Entropie" (signatures les plus stables et les plus complexes). Les mêmes résultats ont été observés dans le Chapitre 7 en confrontant les catégories de personnes générées par la mesure d'Entropie Personnelle aux performances de plusieurs systèmes de vérification soumis à la campagne d'évaluation BSEC'2009 [25, 52]. Nous avons aussi montré dans le Chapitre 7 l'intérêt applicatif potentiel de cette mesure pour différentes applications réelles dans le but d'améliorer les performances des systèmes de vérification.

La deuxième contribution originale de notre travail, présentée dans le Chapitre 5, est une mesure de qualité basée sur le concept d'entropie relative (appelé aussi divergence de Kullback-Leibler). Cette nouvelle mesure, dénommée "Entropie Relative Personnelle" permet de quantifier la vulnérabilité des signatures aux attaques (les bonnes imitations). L'Entropie Relative Personnelle mesure pour une personne donnée la divergence de Kullback-Leibler entre les lois de probabilités locales de ses signatures authentiques et des bonnes imitations qui lui sont associées par le biais de deux Modèles de Markov Cachés, l'un appris sur les signatures authentiques de la personne et l'autre sur les imitations associées à cette personne. Plus la divergence de Kullback-Leibler est faible, plus la personne est considérée comme étant vulnérable aux attaques. Les expériences ont montré que cette nouvelle mesure caractérise les personnes non seulement en termes de complexité, de variabilité et de lisibilité de leur signature, comme notre mesure d'Entropie Personnelle, mais aussi en termes de sa vulnérabilité aux attaques. L'originalité de cette mesure est qu'elle est bien appropriée à l'analyse des systèmes biométriques du fait de sa capacité à trouver un bon compromis entre l'amélioration du FAR et la dégradation du FRR.

La troisième contribution originale, présentée dans le Chapitre 6, propose de quantifier la qualité des signatures imitées, ce qui est totalement nouveau dans la

littérature. Cette mesure est basée sur l'extension de la mesure d'Entropie Personnelle au contexte des imitations. En fait, la qualité d'une imitation est quantifiée par la différence existant entre l'Entropie Personnelle associée à la personne et l'entropie de l'imitation, calculée avec le même Modèle de Markov Caché de cette personne. Cette mesure de qualité des imitations a souligné l'importance des conditions d'acquisition et de l'aspect dynamique pour la production d'une bonne imitation dans le contexte en-ligne. L'impact de cette mesure est important car elle permet lors de l'évaluation des systèmes de vérification de signatures en-ligne, de catégoriser les attaques aux systèmes en différents degrés de qualité, et d'apporter ainsi une information vis-à-vis de la robustesse des systèmes aux attaques.

Les perspectives de ce travail sont nombreuses. Une première perspective consisterait à ***considérer une représentation de signatures plus riche en utilisant d'autres caractéristiques pour le calcul des trois mesures de qualité***. En effet, la plupart des travaux de la littérature abordant l'aspect vérification dans le contexte en-ligne, utilisent des approches de classification combinées à une représentation de la signature plus riche incluant des caractéristiques spatiales et dynamiques. Il serait alors intéressant d'étudier le comportement de nos mesures de qualité confrontées à plusieurs extractions de caractéristiques et d'analyser les catégories de personnes selon les caractéristiques considérées.

Cela nous amène à explorer une autre voie de recherche celle de la ***sélection de caractéristiques de la signature***. En fait, la stabilité, la complexité et la vulnérabilité des signatures sont totalement liées aux caractéristiques considérées pour représenter la signature. Une bonne représentation de la signature est celle qui à la fois caractérise le mieux une personne en termes de complexité et stabilité de ses signatures, et qui aussi discrimine le mieux des imitations les signatures authentiques de cette personne. On propose ainsi d'élaborer une méthode de sélection de caractéristiques basée sur la mesure d'Entropie Relative, qui comme nous l'avons vu dans cette thèse, trouve un bon compromis entre caractérisation de la personne et sa discrimination des attaques. On choisira alors les caractéristiques qui rendent les signatures d'une personnes plus stables, plus complexes, et plus résistantes aux attaques, donc celles qui maximisent l'Entropie Relative.

Une autre direction de recherche est d'exploiter les mesures de qualité introduites dans cette thèse pour ***mesurer la qualité des bases de signatures en-ligne*** pour de futures évaluations internationales de systèmes de vérification. Cela permettrait d'évaluer la qualité des données de la base sans avoir recours à l'étape de classification, et de catégoriser la base en sous-ensembles de données

de qualités différentes. Nous avons proposé cela lors de la compétition internationale BSEC'2009 [25, 52] en évaluant les systèmes par catégories de personnes générées avec l'Entropie Personnelle. Toutefois, on pourrait aussi générer avec la mesure d'Entropie Relative Personnelle des catégories de personnes selon leur vulnérabilité aux attaques, et confronter ces catégories de personnes à des imitations de qualité différentes suivant la mesure de qualité des imitations. Nous avons ainsi en perspective d'organiser une nouvelle campagne d'évaluation internationale de signatures mandatée par l'Association BioSecure [54], à la conférence internationale IJCB'2011 (International Joint Conference on Biometrics). Cette ***nouvelle campagne d'évaluation ESRA'2011*** (Evaluation of Signature Resistance to Attacks [53]) aura pour but d'évaluer la résistance de différents systèmes de vérification à des imitations de différents degrés de qualité générés par notre mesure de qualité des imitations [53].

Un autre axe de recherche intéressant consisterait à ***intégrer la mesure de qualité d'Entropie Personnelle dans la phase de classification*** pour améliorer les performances des systèmes de vérification. En effet, on pourrait détecter les zones les plus stables d'une signature en exploitant l'aspect local de la mesure d'entropie, puis d'octroyer une plus grande importance aux portions stables de la signature qu'aux portions instables lors de la vérification.

Finalement, les trois mesures de qualité proposées dans cette thèse ont été calculées sur des signatures représentées par des caractéristiques spatiales. Nous avons vu dans l'état-de l'art que la qualité des signatures hors-ligne et en-ligne est caractérisée avec les mêmes critères (stabilité, complexité et lisibilité). Il serait alors intéressant d'***étendre les trois mesures de qualité proposées dans cette thèse à la modalité signature hors-ligne*** en se basant sur une estimation locale des densités de probabilité par le biais des Modèles de Markov Cachés, qui sont adaptés naturellement à la modélisation des signatures hors-ligne.

Bibliographie

[1] Andy Adler, Richard Youmaran, and Sergey Loyka. Towards a measure of biometric feature information. *Pattern Analysis and Applications*, 12(3), September 2009.

[2] C. Allgrove and M.C. Fairhurst. Enrolment Model Stability in Static Signature Verification. In *Proc. of International Workshop on Frontiers in Handwriting Recognition (IWFHR'2000)*, pages 565–570, Amsterdam, Netherlands, September 2000.

[3] F. Alonso-Fernandez, M. C. Fairhurst, J. Fierrez, and J. Ortega-Garcia. Automatic measures for predicting performance in off-line signature. In *IEEE Proc. International Conference on Image Processing, ICIP*, volume 1, pages 369–372, San Antonio TX, USA, September 2007.

[4] F. Alonso-Fernandez, M. C. Fairhurst, J. Fierrez, and J. Ortega-Garcia. Impact of signature legibility and signature type in off-line signature verification. In *Biometrics Symposium, BSYM, IEEE*, pages 1–6, Baltimore, USA, September 2007.

[5] F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Gonzalez-Rodriguez. Quality-based conditional processing in multi-biometrics : application to sensor interoperability. *IEEE Trans. on Systems, Man and Cybernetics Part A*, 40(6) :1168–1179, 2010.

[6] L. Ballard, D. Lopresti, and F. Monrose. Forgery Quality and its Implications for Behavioral Biometric Security. *IEEE Transactions on Systems, Man, and Cybernetics-Part B : Cybernetics (Special Edition)*, 37(5) :1107–1118, 2007.

[7] M. Bendris, D. Charlet, and G. Chollet. Introduction of quality measures in audio-visual identity verification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '09)*, pages 996–1000, 1997.

[8] M. C. Bishop. *Pattern Recognition and Machine Learning*. Information Science and statistics. Springer, 2006.

[9] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior. Guide to Biometrics. Springer-Verlag, 2003.

[10] R. Bolle, S. Pankanti, and N. Ratha. Evaluation Techniques for Biometrics-based Authentication Systems (FRR). In *International Conference on Pattern Recognition (ICPR)*, 2000.

[11] V. Boulétreau. *Vers un classement de l'écrit par des méthodes fractales.* PhD thesis, Institut National des Sciences Appliquées de Lyon, Villeurbanne, France, 1997.

[12] V. Boulétreau, N. Vincent, R. Sabourin, and H. Emptozt. Handwriting and signature : one or two personality identifiers ? In *Proc. of the 14th International Conference on Pattern Recognition (ICPR'1998)*, pages 1758–1760, Brisbane, Australia, August 1998.

[13] L. Bovino, S. Impedovo, G. Pirlo, and L. Sarcinella. Multi-Expert Verification of Hand-Written Signature. In *Proc. of the 7th International Conference on Document Analysis and Recognition (ICDAR), IEEE*, 2003.

[14] J. J. Brault and R. Plamondon. How to detect problematic signers for automatic signature verification. In *Proc. of the International Canadian Conference on Security Technology (ICCST)*, pages 127–132, Zurich, Switzerland, 1989.

[15] J. J. Brault and R. Plamondon. A Complexity Measure of Handwritten Curves : Modeling of Dynamic Signature Forgery. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(2) :400–413, 1993.

[16] V. S. Chakravarthy and B. Kompella. The shape of handwritten characters. *Pattern Recognition Letters*, 24(12) :1901–1913, August 2003.

[17] W. D. Chang and J. Shin. Modified Dynamic Time Warping for Stroke-Based On-line Signature Verification. In *Proc. of the 9th International Conference on Document Analysis and Recognition (ICDAR'2007)*, volume 2, pages 724–728, September 2007.

[18] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Second Edition, John Wiley and Sons, 2006.

[19] J. Daugman. Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 36(7) :1169–1179, 1988.

[20] J. Daugman. The importance of being random : statistical principles of iris recognition. *Pattern Recognition*, 36 :279–291, 2002.

[21] G. Dimauro, S. Impedovo, R. Modugno, G. Pirlo, and L. Sarcinella. Analysis of stability in hand-written dynamic signatures. In *Proc. of the 8th International Workshop on Frontiers in Handwriting Recognition (ICFHR'02)*, Ontario, Canada, August 2002.

[22] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheeps, Goats, Lambs and Wolves, A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation. In *International Conference on Spoken Language Processing (ICSLP'98)*, 1998.

[23] J. G. A. Dolfing. *Handwriting recognition and verification, a Hidden Markov approach.* PhD thesis, Philips Electronics N.V., 1998.

[24] J. G. A. Dolfing, E. H. L. Aarts, and J. J. G. M. Van Oosterhout. On-line signature verification with hidden markov models. In *Proc. of the International Conference on Pattern Recognition*, pages 1309–1312, Brisbane, Autralia, 1998.

[25] B. Dorizzi, R. Cappelli, M. Ferrara, D. Maio, D. Maltoni, N. Houmani, S. Garcia-Salicetti, and A. Mayoue. Fingerprint and On-Line Signature Verification Competitions at ICB 2009, LNCS Vol. 5558, Springer. In *International Conference on Biometrics (ICB 2009)*, pages 725–732, June 2009.

[26] S. Dudoit and J. Fridlyand. A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biology*, 3(7) :1–21, 2002.

[27] J.A. Euchner, J.H. Coffy, and A. Obrea. System and Method for Annoting Documents. *US Patent 7111230*, September 2006.

[28] J. Fabregas and M. Faundez-Zanuy. Biometric dispersion matcher. *Pattern Recognition*, 41(11) :3412–3426, 2008.

[29] J. Fabregas and M. Faundez-Zanuy. Biometric dispersion matcher versus LDA. *Pattern Recognition*, 42(9) :1816–1823, 2009.

[30] J. Fabregas and M. Faundez-Zanuy. On-line signature verification system with failure to enrol management. *Pattern Recognition*, 42(9) :2117–2126, 2009.

[31] M.C. Fairhurst and E. Kaplani. Perceptual analysis of handwritten signatures for biometric authentication. *IEE Proc. Vision, Image and Signal Processing*, 150(6) :389–394, 2003.

[32] F. Alonso Fernandez. *Biometric Sample Quality and its Application to Multimodal Authentication Systems.* PhD thesis, Universidad Politecnica de Madrid, October 2008.

[33] J. Fierrez, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez. HMM-based on-line signature verification : feature extraction and signature modeling. *Pattern Recognition Letters*, 28(16) :2325–2334, December 2007.

[34] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Peñalba, J. Ortega-Garcia, and D. Maltoni. An on-line signature verification system based on fusion of local and global information. In *Proc. of 5th IAPR Intlernational Conference on Audio- and Video-based Biometric Person Authentication, AVBPA*, volume 3546 of *LNCS*, pages 407–414, New York, USA, July 2005. Springer.

[35] M. Fuentes, S. Garcia-Salicetti, and B. Dorizzi. On-line Signature Verification : Fusion of a Hidden Markov Model and a Neural Network via a Support Vector Machine. In *Proc. of International Workshop on Frontiers of Handwritten Recognition*, pages 253–258, Niagara on the Lake, Canada, August 2002.

[36] M. Fuentes, D. Mostefa, J. Kharroubi, S. Garcia-Salicetti, B. Dorizzi, and G. Chollet. Identity verification by fusion of biometric data : on-line signature and speech. In *Proc. of the COST-275 Workshop*, Rome, Italy, 2002.

[37] S. Furui. Cepstral Analysis Technique for Automatic Speaker Verification. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 29 :254–272, 1981.

[38] J. Galbally, J. Fierrez, M. R. Freire, and J. Ortega-Garcia. Feature selection based on genetic algorithms for on-line signature verification. In *Proc. IEEE Workshop on Automatic Identification Advanced Technologies, AutoID*, pages 198–203, Alghero, Italy, June 2007.

[39] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. Leroux-Les Jardins, J. Lanter, Y. Ni, and D. Petrovska-Delacretaz. BIOMET : a Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities. In *Proc. of 4th International Conference on Audio and Vidio-Based Biometric Person Authentication*, pages 845–853, Guildford, UK, 2003.

[40] S. Garcia-Salicetti, J. Fierrez-Aguilar, F. Alonso-Fernandez, C. Vielhauer, R. Guest, L. Allano, T. Doan Trung, T. Scheidat, B. Ly Van, J. Dittmann, B. Dorizzi, J. Ortega-Garcia, J. Gonzalez-Rodriguez, M. Bacile di Castiglione, and M. Fairhurst. Biosecure Reference Systems for On-Line Signature Verification : A Study of Complementarity. *Annals of Telecommunications, Special Issue on Multimodal Biometrics*, 2007.

[41] S. Garcia-Salicetti and N. Houmani. *Encyclopedia of Biometrics*, chapter Digitizing Tablet, pages 224–228. Ed.Z. Li Stan, Springer-Verlag, Germany, 2009.

[42] S. Garcia-Salicetti, N. Houmani, and B. Dorizzi. A Client-entropy Measure for On-line Signatures. In *Proc. of IEEE Biometrics Symposium (BSYM)*, Tampa, USA, September 2008.

[43] S. Garcia-Salicetti, N. Houmani, and B. Dorizzi. A Novel Criterion for Writer Enrolment based on a Time-Normalized Signature Sample Entropy Measure. *EURASIP Journal on Advances in Signal Processing, Special issue on recent advances in biometric systems : a signal processing perspective*, 2009(15) :1901–1913, January 2009. doi :10.1155/2009/964746.

[44] S. Garcia-Salicetti, N. Houmani, B. Ly-Van, B. Dorizzi, F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, C. Vielhauer, and T. Scheidat. *Guide to Biometric Reference Systems and Performance Evaluation*, chapter Online Handwritten Signature Verification, pages 125–166. Springer-Verlag, London, 2008. ISBN 978-1-84800-291-3.

[45] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. On Clustering Validation Techniques. *Intelligent Information Systems Journal*, 17 :107–145, 2001.

[46] S. Hangai, S. Yamanaka, and T. Hamamoto. Writer Verification using Altitude and Direction of Pen Movement. In *International Conference on Pattern Recognition*, pages 3483–3486, Barcelona, September 2000.

[47] F. Hao and C. W. Chan. Online signature verification using a new extreme points warping technique. *Pattern Recognition Letters*, 24(16) :2943–2951, 2003.

[48] C. Hook, J. Kempf, and G. Scharfenberg. A Novel Digitizing Pen for the Analysis of Pen Pressure and Inclination in Handwriting Biometrics. In *Biometric Authentication Workshop (BioAW), Lecture Notes in Computer Science (LNCS)*, volume 3087, pages 283–294, Prague, Czech Republic, May 2004.

[49] N. Houmani, S. Garcia-Salicetti, and B. Dorizzi. A Novel Personal Entropy Measure confronted with Online Signature Verification Systems' Performance. In *Proc. of IEEE Second International Conference on Biometrics : Theory, Applications and Systems (BTAS'2008)*, Washington, USA, September 2008.

[50] N. Houmani, S. Garcia-Salicetti, and B. Dorizzi. On assessing the Robustness of Pen Coordinates, Pen Pressure and Pen inclination to Time Variability

with Personal Entropy. In *IEEE Third International Conference on Biometrics : Theory, Applications and Systems (BTAS'2009)*, Washington,USA, September 2009.

[51] http ://biometrics.it sudparis.eu/BMEC2007/.

[52] http ://biometrics.it sudparis.eu/BSEC2009/.

[53] http ://biometrics.it sudparis.eu/ESRA2011/.

[54] http ://biosecure.it sudparis.eu/AB/.

[55] http ://fr.wikipedia.org/wiki/

[56] http ://www.anoto.com/.

[57] http ://www.destinyplc.com/.

[58] http ://www.wacom components.com/english/technology/emr.html.

[59] http ://www.wacom.com/.

[60] L. Hubert and J. Schultz. Quadratic assignment as a general data-analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 1(629) :190–241, 1976.

[61] D. Impedovo and G. Pirlo. Automatic Signature Verification : The State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics-Part C*, 38(5) :609–635, September 2008.

[62] A. K. Jain, F. D. Griess, and S. D. Connell. On-line signature verification. *Pattern Recognition*, 35(12) :2963–2972, December 2002.

[63] A. K. Jain and D. Zongker. Feature Selection : Evaluation, Application, and Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2) :153–158, February 1997.

[64] S. A. Jassim, H. Al-Assam, A. J. Abboud, and H. Sellahewa. Analysis of Relative Entropy, Accuracy, and Quality of Face Biometric. *Pattern Recognition for IT security Workshop*, 2010.

[65] J.Fierrez-Aguilar, D. Ramos-Castro, J. Ortega-Garcia, and J. Gonzales-Rodriguez. HMM-based on-line signature verification : feature extraction and signature modelling. *Pattern Recognition Letters*, 28(16) :2325–2334, December 2007.

[66] K. Jung-Tae and B. Sung-Yang. A Measure of Recognition Difficulty for a Character Image Database. In *International Conference on Document Analysis and Recognition (ICDAR'97)*, pages 996–1000, 1997.

[67] R. Kashi, J. Hu, W.L. Nelson, and W. Turin. A Hidden Markov Model approach to online handwritten signature verification. *International Journal on Document Analysis and Recognition*, 1 :102–109, 1998.

[68] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. *Knowledge and Information Systems*, 3(3) :263–286, 2000.

[69] H. Ketabdar, J. Richiardi, and A. Drygajlo. Global Feature Selection for Online Signature Verification. In *Proc. of International Graphonomics Society 2005 Conference*, pages 625–629, 2005.

[70] A. Kholmatov and B. Yanikoglu. An Individuality Model for Online Signatures Using Global Fourier Descriptors. In *SPIE Defense and Security : Biometric Technology For Human Identification V*, Orlando, USA, March 2008.

[71] A. Kholmatov and B.A. Yanikoglu. An improved decision criterion for genuine/forgery classification in on-line signature verification. In *Proc. of International Conference on Artificial Neural Networks (ICANN)*, Istanbul, Turkey, June 2003.

[72] A. Kholmatov and B.A. Yanikoglu. Identity authentication using improved online signature verification method. *Pattern recognition letters*, 26(15) :2400–2408, November 2005.

[73] R. Kohavi. A study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995.

[74] Y. Komiya and T. Matsumoto. On-line Pen Input Signature Verification PPI (Pen-Position / Pen-Pressure / Pen-Inclination). *IEEE*, 1999.

[75] S. Kullback and R.A. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics*, 22(1) :79–86, 1951. doi :10.1214/aoms/1177729694. MR39968.

[76] V. Di lecce, G. Dimauro, A. Guerriero, S. Impedovo, G. Pirlo, A. salzo, and L. Sarcinella. Selection of Reference Signatures for Automatic Signature Verification. In *the 5th International Conference on Document Analysis and Recognition (ICDAR'99)*, pages 597–600, 1999.

[77] N. Liu and Y.H. Wang. Template selection for on-line signature verification. In *International Conference on Pattern Recognition (ICPR'08)*, pages 1–4, 2008.

[78] D. Maramatsu and T. Matsumoto. Online signature verification algorithm with a user-specific global-parameter fusion model. In *IEEE International Conference on Systems, Man, and Cybenetics*, pages 492–497, 2009.

[79] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The DET Curve in Assessment of Detection Task Performance. In *Proc. of EUROSPEECH*, volume 4, pages 1895–1898, Rhodes, Greece, 1997.

[80] M. Martinez-Diaz, J. Fierrez, J. Galbally, and J. Ortega-Garcia. Towards mobile authentication using dynamic signature verification : useful features and performance evaluation. In *Proc. of the International Conference on Pattern Recognition, ICPR*, Tampa, USA, December 2008.

[81] M. Martinez-Diaz, J. Fierrez, and S. Hangai. *Encyclopedia of Biometrics*, chapter Signature Features. Springer Verlag, July 2009.

[82] M. Martinez-Diaz, J. Fierrez, and S. Hangai. *Signature Features*. Springer Verlag, July 2009.

[83] M. Martinez-Diaz, J. Fierrez, and S. Hangai. *Signature Matching*. Springer Verlag, July 2009.

[84] A. Mellakh. *Reconnaissance du visage en conditions dégradées*. PhD thesis, TELECOM SudParis, France, Avril 2009.

[85] S. Müller and O. Henniger. Evaluating the Biometric Sample Quality of Handwritten Signatures. In *Proc. of the International Conference on biometrics (ICB), Springer LNCS 4642*, pages 407–414, Seoul, South Korea, 2007.

[86] I. K. Mostafa, M. Mohamed, and H. Abbas In. Enhanced DTW based on-line signature verification. In *ICIP*, pages 2713–2716, 2009.

[87] D. Muramatsu and T. Matsumoto. An HMM On-line Signature Verifier Incorporating Signature Trajectories, 2003.

[88] D. Muramatsu and T. Matsumoto. Effectiveness of Pen Pressure, Azimuth, and Altitude Features for Online Signature Verification. In *Proc. of IAPR International Conference on Biometrics, ICB*, pages 503–512, Korea, 2007.

[89] J. P. Nakache and J. Confais. *Approche pragmatique de la classification : Arbres hiérarchiques, Partitionnements*. Editions Technip, 2004.

[90] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M.R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J.L Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M.N.R Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F.M. Sukno, S.K. Pavani, A. Frangi, L. Akarun, and A. Savran. The Multiscenario Multienvironment BioSecure Multimodal Database

(BMDB). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6), June 2010.

[91] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro. MCYT Baseline Corpus : A Bimodal Biometric Database. *IEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, 150(6) :395–401, December 2003.

[92] J. Ortega-Garcia, J. Gonzalez-Rodriguez, D. Simon-Zorita, and S. Cruz-Llanas. From Biometrics Technology to Applications regarding face, voice, signature and fingerprint Recognition Systems. In D. Zhang, editor, *in Biometrics Solutions for Authentication in an E-World*, pages 289–337. Kluwer Academic Publishers, July 2002.

[93] J.M. Pascual-Gaspar. *Uso de la Firma Manuscrita Dinámica para el Reconocimiento Biométrico de Personas en Escenarios Prácticos*. PhD thesis, epartamento de Informática. Universidad de Valladolid, January 2010. http ://uvadoc.uva.es/handle/10324/130.

[94] J.M. Pascual-Gaspar, V. Cardeñoso-Payo, and C.E. Vivaracho-Pascual. Practical on-line signature verification. In *ICB 2009*, pages 1180–1189, 2009.

[95] R. Plamondon and G. Lorette. Automatic signature verification and writer identification : The state of the art. *Pattern Recognition*, 22(2) :107–131, 1989.

[96] L. Rabiner and B.H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall Signal Processing Series, 1993.

[97] D. A. Reynolds and R. C. Rose. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1) :72–83, January 1995.

[98] J. Richiardi. *Probabilistic models for multi-classifier biometric authentication using quality measures*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Suisse, 2007.

[99] J. Richiardi and A. Drygajlo. Gaussian Mixture Models for On-line Signature Verification. In *International Multimedia Conference, Proc. 2003 ACM SIGMM workshop on Biometrics methods and applications*, pages 115–122, Berkeley, California, USA, 2003.

[100] J. Richiardi, H. Ketabdar, and A. Drygajlo. Local and global feature selection for on-line signature verification. In *Proc. of the 8th International Conference on Document Analysis and Recognition (ICDAR'2005)*, 2005.

[101] G. Rigoll and A. Kosmala. A systematic comparison of on-line and off-line methods for signature verification with hidden markov models. In *Proc. of the 14th International Conference on Pattern Recognition*, pages 1755–1757, Brisbane, Autralia, 1998.

[102] O. Rioul. *Théorie de l'information et du codage*. Hermes Science - Lavoisier, 2007.

[103] O. Rohlik. *Encyclopedia of Biometrics*. PhD thesis, University of West Bohemia in Pillsen, Czech republic, March 2003.

[104] R. Sabourin. Off-line signature verification : Recent advances and perspectives. *Lecture notes in Computer Science, 1339*, pages 84–98, 1997.

[105] G. Saporta. *Probabilités, analyses des données et statistiques*. Editions Technip, deuxième édition, 2006.

[106] Y. Sato and K. Kogure. Online signature verification based on shape, motion and writing pressure. In *Proc. of the 16th International Conference on Pattern Recognition*, pages 823–826, 1982.

[107] A. Strehl. *Relationship-based Clustering and Cluster Ensembles for High-dimensional Data Mining*. PhD thesis, University of Texas at Austin, May 2002.

[108] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 2003.

[109] B. Ly Van. *Réalisation d'un Système de Vérification de Signature Manuscrite En-ligne Indépendant de la Plateforme d'Acquisition*. PhD thesis, Institut Telecom ; Telecom SudParis, 2005.

[110] B. Ly Van, S. Garcia-Salicetti, and B. Dorizzi. On using the Viterbi Path along with HMM Likelihood Information for On-line Signature Verification. *IEEE Transactions on Systems, Man and Cybernetics, Part B, Special Issue on Recent Advances in Biometric Systems*, 37(5) :1237–1247, October 2007.

[111] C. Vielhauer. *Biometric User Authentication for IT Security : From Fundamentals to Handwriting*. Springer, New York, USA, 2006.

[112] C. Vielhauer, R. Steinmetz, and A. Mayerhöfer. Biometric Hash based on Statistical Features of Online Signature. In *International Conference on Pattern Recognition*, 2002.

[113] X. Wang, X. Ding, and H. Liu. Writer Identification Using Directional Element Features and Linear Transform. In *Proc. of the 7th International Conference on Document Analysis and Recognition*, volume 2, 2003.

[114] M. Wirotius, J.Y. Ramel, and N. Vincent. Selection of Points for On-Line Signature Comparison. In *International Workshop On Frontiers in Handwriting Recognition (IWFHR)*, pages 503–508, 2004.

[115] M. Wirotius, J.Y. Ramel, and N. Vincent. Contribution of global temporal information for authentication by on-line handwritten signatures. In *Actes de IGS'05*, pages 266–270, 2005.

[116] M. Wittman, P. Davis, and P.J. Flynn. Empirical Studies of the Existence of the Biometric Menagerie in the FRGC 2.0 Color Image Corpus. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, 2006.

[117] N. Yager and T. Dunstone. Worms, Chameleons, Phantoms and Doves : New Additions to the Biometric Menagerie. In *IEEE Workshop on Automatic Identification Advanced Technologies*, 2007.

[118] N. Yager and T. Dunstone. The biometric menagerie. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(2), 2010.

[119] Y. Yamazaki, T. Nagao, and N. Komatsu. Text-indecated Writer Verification Using Hidden Markov Models. In *In Proc. of the 7th International Conference on Document Analysis and Recognition (ICDAR)*, 2003.

[120] L. Yang, B.K Widjaja, and R. Prasad. Application of Hidden Markov Models for Signature Verification. *Pattern Recognition*, 28(2) :161–170, 1995.

[121] B.A. Yanikoglu and A. Kholmatov. Online Signature Verification Using Fourier Descriptors. *EURASIP Journal on Advances in Signal Processing*, 2009. Article ID 260516, doi :10.1155/2009/260516.

[122] D. Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll. SVC2004 : First International Signature Verification Competition. In *International Conference on Biometric Authentication (ICBA), LNCS*, volume 3072, pages 16–22, Hong Kong, China, July 2004.

Annexe A

Nombre optimal de catégories de personnes

Cette annexe présente la procédure suivie pour l'évaluation du nombre approprié de catégories de personnes k dans une base de signatures donnée, en l'occurrence MCYT-100 dans ce cas. Cette étude nécessite différents indices de validité de classification [45], considérés comme des outils dont le but est d'évaluer quantitativement les résultats des algorithmes de classification.

Comme nous l'avons vu dans le Chapitre 5, la mesure d'Entropie Relative Personnelle catégorise les personnes en termes de complexité et de variabilité de leur signature comme l'Entropie Personnelle, mais aussi en termes de vulnérabilité de leur signature aux attaques. Nous considérons alors l'Entropie Relative Personnelle comme le meilleur critère pour la catégorisation des personnes. C'est pourquoi dans cette annexe nous avons fait le choix de montrer l'étude du nombre optimal des catégories de personnes sur les valeurs d'Entropie Relative Personnelle seulement. Nous avons suivi la même démarche pour trouver d'une part le nombre optimal de catégories de personnes en considérant leurs valeurs d'Entropie Personnelle dans le Chapitre 4, et d'autre part le nombre optimal de catégories d'imitations par personne dans le Chapitre 6

Pour générer les catégories de personnes suivant leurs valeurs d'Entropie Relative Personnelle, nous avons appliqué une classification hiérarchique ascendante [89] sur ces valeurs en utilisant le critère de "Ward" pour l'agrégation [89]. Le critère de "Ward" consiste à choisir à chaque étape de la classification le regroupement de classes tel que l'augmentation de l'inertie intra-classe I, utilisée comme indice de niveau, soit la plus petite possible pour que les classes restent homogènes.

L'expression de ce critère est :

$$I = \frac{n_a.n_b}{(n_a + n_b)}.d^2(g_a, g_b) \tag{A.1}$$

Où $d^2(g_a, g_b)$ est la distance euclidienne entre les centres de gravité g_a et g_b des deux classes a et b (qui, à une étape donnée, peuvent ne comporter qu'un élément) ; n_a et n_b désignent la somme des poids des éléments de chaque classe (qui, à une étape donnée, peut ne comporter qu'un élément).

A.1 Procédure d'identification du nombre optimal de catégories

A chaque étape de la classification hiérarchique [89], c.à.d. à chaque niveau d'agrégation, on calcule différents indices de validité de la littérature [45] : C-index [60], Weighted intra-inter index [107], Krzanowski Laï index [26], et les 4 indices de validité du RMSSTD Group [89] : Root Mean Square Standard Deviation, R-Squared, semi-partial R-Squared index et la distance entre deux classes.

En se servant de ces indices de validité [45], la procédure suivie pour identifier le nombre optimal de catégories de personnes dans MCYT-100 est la suivante :

1. Appliquer la classification hiérarchique ascendante sur les 100 valeurs d'Entropie Relative associées aux personnes de MCYT-100.

2. Tracer l'arbre hiérarchique (ou *dendrogramme*) affichant l'évolution de l'agrégation (inertie intra-classe) selon le regroupement effectué à chaque étape. La Figure A.1 montre l'arbre hiérarchique issu de la classification hiérarchique ascendante sur les 100 valeurs d'Entropie Relative Personnelle.

3. En déduire un intervalle de variation du nombre de catégories k (définition de k_{min} et k_{max}) en observant les changements importants de l'agrégation dans le dendrogramme : en partant des faibles valeurs d'agrégations, on cherche les premières grandes distances d'agrégation c.à.d. les premiers grands écarts en hauteur entre deux nœuds successifs ; puis en chaque palier correspondant à ces écarts de distance on coupe l'arbre hiérarchique par une droite horizontale. Toute coupure d'une droite horizontale avec l'arbre fournit alors une partition possible. Comme l'illustre la Figure A.1, les changements importants de l'agrégation sont représentés par 4 lignes horizontales discontinues correspondant à une partition de la base

MCYT-100 en 2, 3, 5 ou 9 catégories de personnes. Par conséquent, dans la suite, nous limitons la zone de recherche du nombre optimal de catégories entre 2 et 10 catégories au maximum.

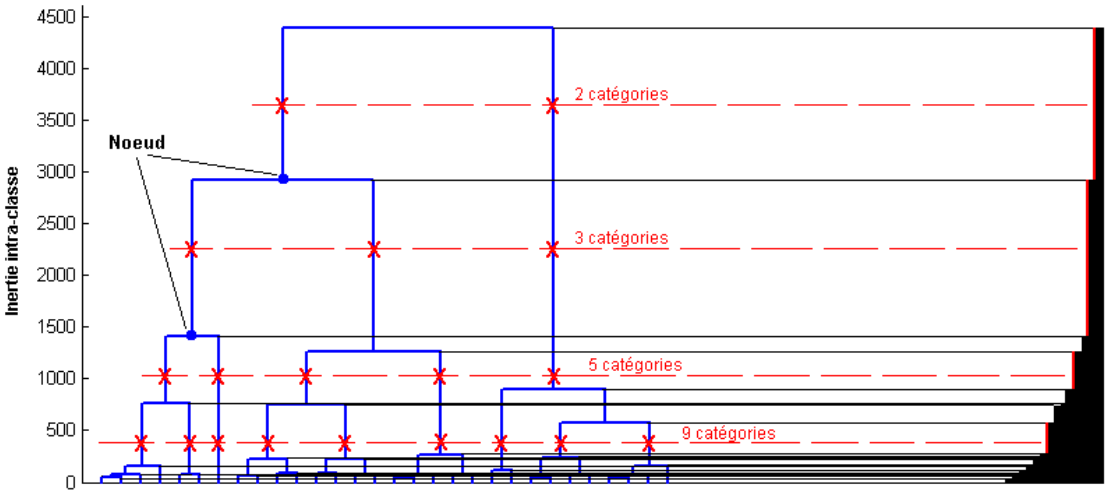

FIGURE A.1: Arbre hiérarchique issu de la classification des personnes de MCYT-100 suivant leur valeur d'Entropie Relative Personnelle. A droite, la projection des niveaux de nœuds.

4. Pour chaque valeur de k entre $k_{min} = 2$ et $k_{max} = 10$, on calcule les différents indices de validité évoqués précédemment.

5. Pour chaque indice de validité, on trace la valeur obtenue de l'indice en fonction de k. La Figure A.2 montre les graphiques des indices de validité calculés pour chaque valeur de k.

6. En se basant sur le tracé obtenu, le nombre optimal de catégories est identifié suivant deux approches [45] : si l'indice de validité en fonction de k est une fonction monotone, le nombre optimal de catégories correspond à la valeur de k où une importante variation locale (augmentation ou diminution) est observée ; sinon, le nombre optimal de catégories est identifié suivant le critère intrinsèque de chaque indice :

- C-index [60] doit être minimisé ;
- Weighted intra-inter index [107] doit être maximisé ;
- Krzanowski Laï index [26] doit être maximisé ;
- Les 4 indices de validité du RMSSTD group [89] doivent être utilisés simultanément, et le nombre optimal de catégories correspond à la valeur de k où une variation locale significative de ces 4 indices est observée.

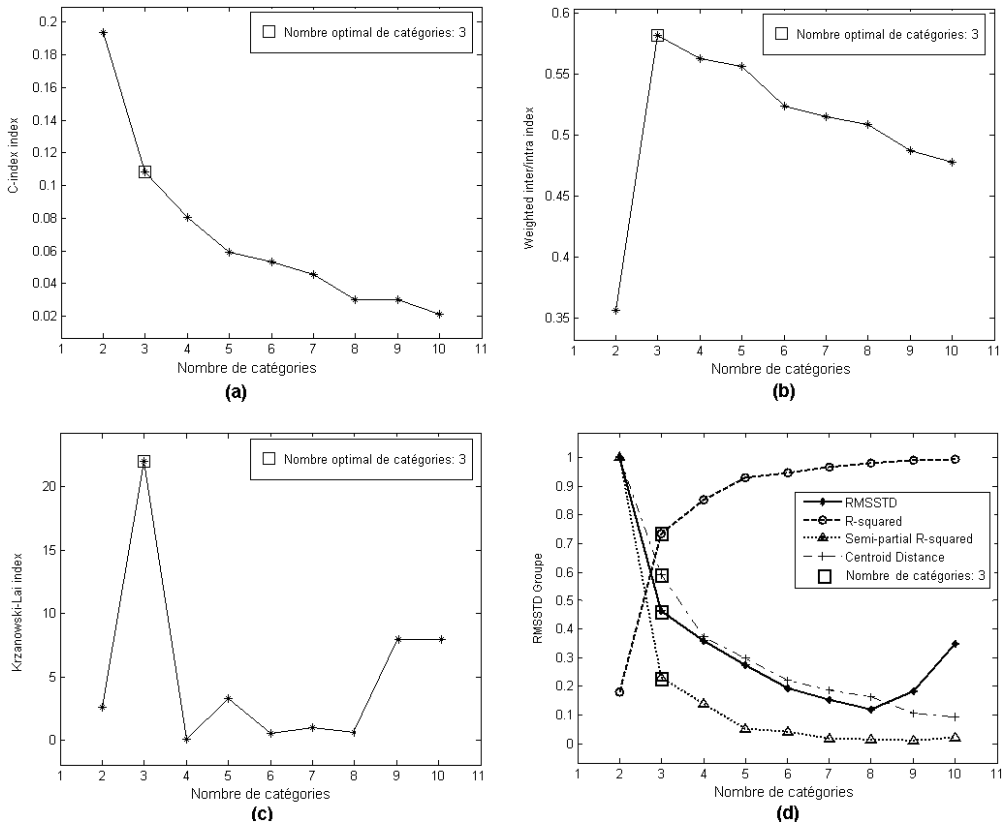

FIGURE A.2: Indices de validité calculés pour chaque valeur de nombre de catégories k : (a) C-index, (b) Weighted intra-inter index, (c) Krzanowski-Lai index, et (d) RMSSTD Group.

Sur la Figure A.2, le nombre optimal de catégories pour chaque indice est indiqué par un symbole carré sur la courbe de l'indice. On constate que tous les indices de validité indiquent que le nombre optimal de catégories est $k = 3$.

En effet, suivant l'étape 6 de la procédure décrite ci-dessus, on observe sur la Figure A.2.a que la variation du C-index en fonction de k est monotone, ainsi le premier coude (variation locale) correspond au nombre optimal de catégories qui est de 3. En ce qui concerne les indices Weighted intra-inter (Figure A.2.b) et Krzanowski-Lai (Figure A.2.c), leur variation en fonction de k n'est pas monotone. Ainsi, le nombre optimal de catégories est défini par la valeur maximale de ces indices (qui est de 3), selon leur critère d'optimalité. De même pour le RMSSTD group (Figure A.2.d), le nombre optimal de catégories est indiqué par la première variation des 4 indices simultanément.

Ainsi, on conclut que $k = 3$ est le nombre optimal de catégories de personnes dans MCYT-100 en considérant leurs valeurs d'Entropie Relative.

Annexe B

Campagne d'évaluation BSEC'2009

Récemment, une nouvelle compétition internationale de vérification de signatures manuscrites en-ligne (BSEC'2009 - BioSecure Signature Evaluation Campaign) a eu lieu en 2009, organisée par notre équipe Intermedia à Telecom sudParis. Dans ce qui suit, nous présenterons le contexte de l'évaluation BSEC'2009 et ses enjeux, ainsi que les résultats qui en découlent.

B.1 Motivation et objectifs

Lors des précédentes compétitions internationales, en l'occurrence la première Compétition Internationale de Vérification de Signatures en-ligne (SVC'2004) [122] et la Campagne d'Evaluation Multimodale de BioSecure (BMEC'2007) [51], les données signatures provenaient d'un seul capteur : une tablette graphique dans le cas de SVC'2004 [122] ; et un PDA dans le cas de BMEC'2007 [51]. D'autre part, les performances ont toujours été évaluées globalement sur toute la base des signatures, et ce indépendamment de toute mesure de qualité. De plus, l'impact de la variabilité temporelle n'a pas été étudié dans les deux compétitions.

La campagne d'évaluation BSEC'2009 [52] comprend plusieurs tâches avec des objectifs distincts, en considérant deux bases de signatures en-ligne acquises dans des conditions bien différentes mais contenant les mêmes personnes. La première base de données est la base BioSecure DS3 [54, 90], dont les signatures ont été acquises sur une plateforme mobile (PDA) ; la deuxième base de données

est la base BioSecure DS2 [54, 90], dont les signatures ont été acquises sur une plateforme fixe (tablette graphique). L'objectif de ces deux scénarios d'acquisition est de mesurer l'impact réel des conditions d'acquisition mobiles sur les performances des algorithmes, en utilisant les deux plus grandes bases de signatures en-ligne existantes contenant les mêmes personnes (BioSecure DS2 et DS3 [54, 90]).

Par ailleurs, le problème de la variabilité temporelle des signatures a longuement été étudié dans la littérature, mais à chaque fois sur différentes bases de signatures, avec différents classifieurs et extractions de caractéristiques. Ainsi, BSEC'2009 a pour deuxième objectif l'étude de l'impact de la variabilité temporelle sur les performances des systèmes de vérification et l'évaluation de la pertinence relative au cours du temps des fonctions temporelles capturées par le capteur [52]. Les bases BioSecure DS2 et DS3 [54, 90] sont en effet bien adaptées à l'étude de cette problématique puisqu'elles ont été collectées en deux sessions espacées dans le temps de plusieurs semaines (se référer à la description de ces deux bases au paragraphe 2.4.4).

De surcroît, la mesure d'Entropie Personnelle a été validée dans nos travaux de recherche en termes de performances de deux systèmes de vérification développés par notre équipe, l'un basé sur un Modèle de Markov Caché (MMC) et l'autre sur la distance élastique (DTW) [43, 49, 42]. Il nous a paru intéressant, à ce niveau, de présenter à la communauté scientifique le concept de catégories de personnes générées par la mesure d'Entropie Personnelle, et de les valider en termes de performances de plusieurs systèmes de vérification en utilisant la même base de signatures et le même protocole d'évaluation. Ainsi, le troisième et principal objectif de BSEC'2009 [52] est d'évaluer les performances de différents algorithmes en fonction du contenu d'information dans les signatures en utilisant notre mesure d'Entropie Personnelle.

B.2 Les participants

Après l'annonce de BSEC'2009, 9 universités de différents pays ont montré leur intérêt à cette campagne d'évaluation. Certaines équipes ont participé aux trois tâches avec un ou plusieurs systèmes, d'autres qu'à la deuxième tâche. Finalement, 14 systèmes ont été soumis.

Le Tableau B.1 indique les équipes qui ont participé à BSEC'2009, ainsi que les tâches auxquelles elles ont participé.

ID	Université	Participants	Tâches
1	Escola Universitaria Politecnica de Mataro, Espagne	J.Roure Alcobé	1-2-3
2		J. Fabregas, M. Faundez-Zany	
3	Institut. de recherche U1, Hongrie	Z.T. Kardkovàcs	1-2-3
4	Univ. de Seikei, Japan	D. Maramatsu	1-2-3
5	Univ. Ain Shams, Egypte	M.I. Khalil, M. Mostafa, H. Abbas	1-2-3
6	Univ. de Valladolid, Espagne	J. M. Pascual-Gaspar, V. Cardeñoso-Payo, C. Vivaracho-Pascual	1-2-3
7	Univ. de Sabanci, Turquie	A. Kholmatov, B. Yanikoglu	1-2-3
8	Univ. Autonoma de Madrid, Espagne	M. Martinez-Diaz, J. Fierrez, J. Ortega-Garcia	1-2-3
9			
10			
11			
12			
13	Univ. de Waseda, Japon	T. Matsumoto	2
14	Univ. de Magdebourg, Allemagne	M. Biermann, T. Scheidat	2
Ref	Telecom SudParis, France	Système de référence	1-2-3

TABLEAU B.1: Liste des équipes participant à la campagne d'évaluation BSEC'2009.

B.3 Description des systèmes

B.3.1 Système 1 (Escola Universitaria Politecnica de Mataro, Espagne)

Ce système est inspiré de celui décrit dans [68] appliqué au domaine des bases de données à grandes séquences temporelles. Ce système est basé sur une extraction locale des paramètres en chaque point :

- la longueur de la signature
- les coordonnées x et y
- les dérivées de x et y : $dx_t = x_{t+1} - x_t$ et $dy_t = y_{t+1} - y_t$
- l'accélération en x et y : $ax_t = dx_{t+1} - dx_t$
- la vitesse absolue : $v_t = \sqrt{(dx_t^2 + dy_t^2)}$
- l'accélération absolue : $dv_t = v_{t+1} - v_t$
- le produit des coordonnées x et y : $xy_t = x_t * y_t$
- l'accélération normale : $na_t = ux_t * ay(t) + uy_t * ax_t$, où $ux_t = dx_t/v_t$, $ux_t = 0$ quand $v_t = 0$
- l'accélération tangentielle : $ta_t = ux_t * ax_t + uy_t * ay_t$
- la position du maxima en x et y. Le point x à l'instant t est considéré maximum quand $a_t > a_{t-1} > a_{t-2}$ et $a_t > a_{t+1} > a_{t+2}$
- la position du manima en x et y. Le point x à l'instant t est considéré minimum quand $a_t < a_{t-1} < a_{t-2}$ and $a_t < a_{t+1} < a_{t+2}$

Cependant, la mise en correspondance entre deux signatures se fait par portion. En fait, la signature est découpée en 8 portions égales. Lorsque la longueur de la signature n'est pas un multiple de 8, les points à la frontière des portions

sont partagés entre ces portions. Puis, en tenant compte de l'ordre des portions, la moyenne M_p et l'écart type E_p de chaque paramètre sont calculés sur chaque portion. Ainsi, chaque signature est représentée par deux vecteurs comprenant respectivement les valeurs moyenne M_p et les valeurs écart-type E_p associées à chaque portion de la signature.

La similarité entre deux signatures R et T est calculée en comparant leurs vecteurs moyenne et écart-type comme suit :

$$S(R,T) = median(ratio(R,T)) \tag{B.1}$$

où *ratio(R, T)* est un vecteur à composantes ratio des valeurs moyenne et écart-type de R et T. Pour que ce ratio soit symétrique, il est calculé comme suit :

$$ratio(R,T) = \begin{cases} \dfrac{T}{R} & si|R| > |T| \\ \dfrac{R}{T} & si|R| < |T| \end{cases} \tag{B.2}$$

A noter que la longueur de la signature est traitée comme un vecteur à une composante. Le score de similarité final entre deux signatures est la somme des différentes similarités des différents paramètres.

B.3.2 Système 2 (Escola Universitaria Politecnica de Mataro, Espagne)

Pour l'extraction des paramètres, ce système utilise une méthode spectrale avec une Transformée en cosinus discrète à une dimension (DCT, ou *Discrete Cosine Transform*) appliquant un filtrage à basse fréquence des variables temporelles [30]. Cette méthode caractérise le comportement dynamique et global de la signature en tenant compte des premiers termes de la DCT [30].

Par ailleurs, une sélection de paramètres est effectuée basée sur un critère de discrimination : l'Analyse Discriminante Linéaire (LDA) [29].

Pour la vérification, ce système utilise un classifieur basé sur l'algorithme de dispersion biométrique [28].

B.3.3 Système 3 (institut de recherche U1, Hongrie)

Ce système est basé sur la fusion par une somme pondérée des scores de 3 systèmes locaux basées sur la distance élastique, et un système global basé sur la comparaison directe des paramètres.

Le 1er système DTW utilise seulement les coordonnées y en entrée. Le score considéré est la moyenne des distances DTW entre la signature de test et les sinatures de référence, normalisé par la moyenne des distances DTW calculées entre toutes les signatures de référence.

Le 2ème système DTW utilise seulement les coordonnées y en entrée. Le score considéré est le minimum des distances DTW entre la signature de test et les signatures de référence, normalisé par le minimum des distances DTW calculées entre toutes les signatures de référence.

Le 3ème système DTW utilise seulement les coordonnées x en entrée. Le score considéré est la moyenne des distances DTW entre la signature de test et les signatures de référence, normalisé par la moyenne des distances DTW calculées entre toutes les signatures de référence.

Le système global utilise 8 paramètres extraits globalement sur toute la signature :

- moyenne des valeurs des coordonnées *(x,y)*, de la pression et les angles d'inclinaison
- écart-types des coordonnées *(x,y)* et de l'angle d'inclinaison *Azimuth*
- nombre des valeurs répétées sur la signature pour les paramètres x, y et l'angle *Azimuth*
- longueur de la signature normalisée dans les directions x et y
- hauteur et largeur de la signature
- moyenne de la tangente positive des paramètres x et y
- vitesse normalisée du stylo en x et y
- nombre de point total dans la signature

Le score de similarité entre la signature de test et les signatures de référence est calculé en comparant par une simple distance les paramètres de la signature de test à ceux de la base de référence. Si la différence se trouve à l'intérieur d'une bande de tolérance prédéfinie, le système attribue une valeur prédéfinie. Le score de similarité de ce système global correspond alors au total de ces valeurs calculées pour les 8 paramètres.

B.3.4 Système 4 (Univ. de Seikei, Japon)

Ce système [78] utilise un classifieur basé sur la distance élastique DTW. Quatre paramètres sont extraits localement sur la signature : les coordonnées x et y, la direction et la vitesse du stylo.

Basé sur la distance élastique (DTW), un score de dissimilarité à 4 dimensions $D = (d_i)$, $i=1,...,4$ est calculé à partir des 4 paramètres entre la signature de test et chacune des signatures de référence m. Le score de décision final est calculé comme suit :

$$Score = \frac{1}{M} \sum_{m=1}^{M} F(D, Mean_{ID}; \Theta) \tag{B.3}$$

où, $F(.;\Theta)$ est un modèle de fusion avec un paramètre Θ, généré en combinant plusieurs perceptrons simples en utilisant l'algorithme d'AdaBoost. $Mean_{ID}$ est le vecteur moyenne à 4 dimensions associé à l'utilisateur ID calculé en utilisant les scores de dissimilarité DTW entre les M signatures de référence associées à cet utilisateur.

B.3.5 Système 5 (Univ. de Ain Shams, Egypte)

Ce système [86] utilise un classifieur basé sur la distance élastique DTW. Deux paramètres extraits localement sont utilisés : la vitesse et les changements de direction définis par le sinus de l'angle entre deux points successifs.

Pour chaque paramètre p, le système calcule la distance DTW entre chaque paire de signatures de référence. Ainsi, à chaque signature de référence lui est associées 4 distances élastiques. Parmi ces 4 distances, seulement la plus faible et la plus grande sont considérées. Par la suite, ces faibles et grandes distances associées aux 5 signatures de référence sont moyennées pour obtenir $avg(min_p)$ et $avg(max_p)$ respectivement.

Le score de dissimilarité associé à la signature de test est calculé comme suit :

$$score = \sum_{p} \left[\frac{min_p(DTW(x,t)) - avg(min_p)}{avg(min_p)} + \frac{max_p(DTW(x,t)) - avg(max_p)}{avg_(max_p)} \right] \tag{B.4}$$

où $min_p(DTW(x,t))$ et $max_p(DTW(x,t))$ représentent la distance minimale et maximale respectivement entre la signature de test et les signatures de référence considérant le paramètre p.

B.3.6 Système 6 (Univ. de Valladolid, Espagne)

Ce système utilise un classifieur basé sur la distance élastique DTW incluant une technique de normalisation du score utilisée spécialement pour cette compétition.

Les paramètres de la signature utilisés sont les dérivées temporelles des coordonnées (x, y). Les paramètres pression et angles d'inclinaison n'ont pas été considérés car ils ont montré une dégradation des performances du système. Une Z-norm statistique est effectuée sur les deux paramètres lors du prétraitement.

Le score de dissimilarité calculé correspond à la distance DTW minimale entre la signature de test et les signatures de référence. Ce système a déjà montré d'excellents résultats avec un seuil de décision personnalisé [94].

Pour adapter ce système aux conditions de la compétition où un seuil de décision commun à toutes les personnes est utilisé, une nouvelle méthode de normalisation a été introduite dénommée *EER-norm* :

$$s_n = \frac{s - \alpha \cdot s_{EER}(C, I)}{\beta \cdot (s_{max}(C) - s_{min}(C)) + (1 - \beta) \cdot (s_{max}(I) - s_{min}(I))} \tag{B.5}$$

Cette technique de normalisation du score requiert deux distributions de score : $s(C)$, obtenue à partir des signatures de référence ; $s(I)$, dérivée du cohorte des imposteurs sélectionnés aléatoirement à partir de la base MCYT-100 [91] dont les caractéristiques sont similaires à celles de la base de la compétition.

Les valeurs de α et β dans l'équation B.5 sont obtenues empiriquement sur la base de développement et optimisés séparément pour chacune des deux bases DS2 et DS3. La valeur de $s_{EER}(C, I)$ représente une estimation '*a priori*' du seuil à erreur égale calculée par la régression multivariée linéaire sur les 16 paramètres extraits des distributions $s(C)$ et $s(I)$ comme décrit dans [93].

B.3.7 Système 7 (Univ. de Sabanci, Turquie)

Ce système est une version modifiée de celui qui a remporté la compétition internationale SVC'2004 [122], décrit dans [72]. Il utilise un classifieur basé sur la distance élastique DTW.

Aucun prétraitement n'est effectué sur les signatures. Les paramètres locaux utilisés sont les différences relatives *(dx,dy)* entre les points de la signature.

Pour la vérification, le système calcule les 3 distances DTW *min, max,* et *moyenne* résultant de la comparaison entre la signature de test et toutes les signatures de référence, normalisées par leurs valeurs respectives obtenues sur la base de référence. Le score final est obtenu à partir de ces 3 distances normalisées en utilisant une Analyse en Composantes Principales (PCA ou *Principal Component Analysis*).

B.3.8 Systèmes 8, 9, 10, 11 et 12 (Univ. Autonoma de Madrid, Espagne)

Cinq systèmes ont été soumis pour toutes les tâches. Ces systèmes utilisent différents paramètres en entrée et différents classifieurs. Une sélection de paramètres par la méthode "Sequential Forward Floating Selection" est effectuée pour adapter chaque système à chaque tâche.

B.3.8.1 Systèmes 8 et 9

Ces deux systèmes utilisent tous les deux un classifieur basé sur la distance élastique [83] mais diffèrent dans la technique de normalisation du score utilisée et dans les paramètres considérés en entrée. Les paramètres sont sélectionnés par la méthode "Sequential Forward Floating Selection" à partir d'un ensemble de 27 paramètres au départ, extraits localement sur la signature. Cet ensemble de paramètre est décrit dans [81]. Le critère d'optimisation est la valeur du *EER* du système sur les bonnes imitations. Le processus de sélection des paramètres est effectué séparément sur DS2 et DS3. Ainsi les paramètres utilisés en entrée des systèmes sur DS3 diffèrent de ceux utilisés en entrée des systèmes sur DS2.

Le premier système (système 8) est optimisé sur les imitations aléatoires via la sélection des paramètres. Pour la base DS3, le score considéré correspond à la distance minimale entre la signature de test et les signatures de référence. Pour la base DS2, le score considéré correspond à la distance moyenne entre la signature de test et les signatures de référence.

Le deuxième système (système 9) est optimisé sur les bonnes imitations via la sélection des paramètres. Une normalisation du score personnalisée est effectuée comme décrit dans [72]. Cette normalisation améliore les performances sur les bonnes imitations mais pénalise les performances sur les imitations aléatoires.

B.3.8.2 Système 10

Ce système, basé sur celui décrit dans [65], utilise un classifieur statistique basé sur les Modèles de Markov Cachés. Un ensemble de 27 paramètres locaux est considéré, le même que celui utilisé pour les systèmes 8 et 9. Une sélection de paramètres est effectuée à partir de cet ensemble par la méthode "Sequential Forward Floating Selection".

B.3.8.3 Système 11

Ce système est basé sur celui décrit en [80] utilisant un vecteur de 100 paramètres définis en [34]. Le score de similarité utilisé est l'inverse de la distance de Mahalanobis.

B.3.8.4 Système 12

Ce système est basé sur la fusion des scores des systèmes 8, 9, 10 et 11 par une moyenne pondérée. Les coefficients de la somme optimaux sont calculés par une régression logistique [5].

B.3.9 Système 13 (Univ. Wesada, Japon)

Dans ce système 7 paramètres ont été utilisés : les coordonnées *(x,y)*, la pression, les angles d'inclinaison, la vitesse absolue et l'angle entre les vecteurs vitesse en x et en y.

Ce système comporte trois phases : phase d'enregistrement, phase d'apprentissage et phase de vérification.

A la phase d'apprentissage, pour chaque paramètre la distance élastique DTW est calculée entre toutes les signatures de test et les signatures de référence. Ainsi, pour chaque personne i, un vecteur $Dist_m(i)$ à 7 dimensions est associé à chaque signature de référence m. On calcule aussi pour chaque paramètre la distance DTW moyenne entre tous les couples possibles des signatures de référence. Ainsi, à chaque personne i est associé un vecteur $Mean_i$ à 7 dimensions. Puis, on considère le vecteur modèle $U_m(i)$ formé par les deux précédents vecteurs comme suit : $U_m(i) = (Dist_m(i), Mean_i)$.

Par la suite, les personnes sont divisées en K groupes en appliquant un K-moyenne sur les vecteurs de moyenne $Mean_i$. Puis, en se basant sur le vecteur modèle $U_m(i)$, les paramètres des K modèles sont estimés à l'aide de l'algorithme Adaboost.

A la phase d'enregistrement, on ne considère que les signatures de référence pour toutes les personnes. La distance DTW entre toutes les signatures de référence est calculée ainsi que le vecteur moyenne permettant ainsi de calculer le vecteur modèle $U_{ref}(i)$. Ce vecteur modèle est après utilisé pour le calcul d'un "score de référence" $(score_k(U_{ref}i))$ permettant de calculer une valeur de fiabilité de chaque modèle k généré à la phase d'apprentissage.

A la phase de vérification, après prétraitement et extraction des paramètres sur la signature de test t, le vecteur modèle $U_t(i)$ est calculé. Le score final est une somme de tous les scores des modèles $(score_k(U_t(i))$ pondérés par leur valeur de fiabilité.

B.3.10 Système 14 (Univ. de Magdebourg, Allemagne)

Ce système est basé sur l'algorithme de Hachage biométrique introduit par *Veilhauer et al.* [112, 111]. Cet algorithme comporte un processus d'enregistrement et un processus de génération de hachage.

Le but du processus d'enregistrement est de générer une matrice d'intervalle *IM* de dimension *N*2* pour chaque personne en considérant N paramètres. Durant ce processus, on calcule pour chaque personne un vecteur D contenant des paramètres statistiques globaux calculés à partir des 5 fonctions temporelles *(x,y,p,Az,Alt)*.

La matrice *IM* contient pour chaque paramètre la largeur d'un intervalle et un offset calculés en se basant sur la variabilité intra-classe de la personne : la largeur de l'intervalle correspond à la différence entre la valeur maximale et la valeur minimale du paramètre considéré sur toutes les signatures de référence ; l'offset correspond au reste de la division euclidienne de la valeur minimale du paramètre considéré par la différence calculée précédemment. Par ailleurs, pour paramétrer la génération de hachage, une valeur de tolérance est déterminée pour chaque paramètre globalement en considérant tous les exemples de signatures disponibles dans la base de développement. Ce système utilise aussi un facteur de tolérance (vecteur scalaire) comme un paramètre de génération de hachage global.

En se basant sur le vecteur de paramètres et sur la matrice d'intervalle associés à la personne ID, le processus de génération de hachage calcule un vecteur de hachage biométrique b_{ID}, où les largeurs intervalle et les offsets de la matrice IM_{ID} sont utilisés pour faire correspondre chaque paramètre statistique k à une valeur de hachage. Pour la vérification, le vecteur de hachage biométrique généré sur la signature test est comparé à celui généré sur la base de référence en utilisant la distance Canberra.

B.3.11 Système de référence (Institut Télécom, France)

Le système de référence est décrit en détail dans [110]. Il utilise 25 paramètres locaux extraits des 5 fonctions temporelles (x, y, p, Az, Alt). Ce système a participé à la compétition SVC'2004 [122] et a été classé $6^{\text{ème}}$ parmi 15 participants dans la Tâche 1 et $2^{\text{ème}}$ parmi 12 participants dans la Tâche 2 [122]. Il a aussi été utilisé comme système de référence à BMEC'2007.

Ce système est basé sur les modèles de Markov Cachés et exploite deux niveaux de description du tracé de la signature. Le score final est la fusion par une simple moyenne arithmétique de deux scores : un score de vraisemblance (information fine moyennée) et un score de segmentation calculé avec l'algorithme de Viterbi (information plus grossière au niveau des portions de la signature).

B.4 Bases de développement et de test

Deux ensembles de développement contenant les signatures de 50 personnes appartenant aux deux bases BioSecure DS2 et DS3 ont été distribués aux participants pour développer leurs systèmes avant soumission. Ces deux ensembles, dénommés DS2-50 et DS3-50 contiennent les mêmes 50 personnes.

Après la soumission des systèmes, ces derniers sont évalués sur deux ensembles de tests dénommés DS2-382 et DS3-382, contenant les signatures de 382 personnes de BioSecure DS2 et DS3 respectivement. A noter, que ces deux ensembles de test sont tenus séquestrés et contiennent les mêmes 382 personnes.

B.5 Protocoles et tâches

B.5.1 Protocole général

Pour l'étude des performances des différents systèmes soumis, le protocol général utilisé est le suivant : pour chaque personne (de DS2 ou de DS3), 5 signatures authentiques de la Session 1 sont utilisées comme signatures de référence. Les tests sont effectués sur les 10 signatures authentiques restantes de la Session 1, les 10 signatures authentiques de la Session 2, les 20 bonnes imitations appartenant aux deux sessions, et sur 30 imitations aléatoires. A noter que selon les tâches, lorsque les tests se font sur les signatures appartenant à la Session 1, il n'y a pas de variabilité temporelle ; lorsque les tests se font sur la Session 2, nous sommes en présence de variabilité temporelle.

B.5.2 Description des différentes tâches d'évaluation

Suivant les objectifs de BSEC'2009 mentionnés précédemment, l'évaluation des différents systèmes a été effectuée en trois tâches :

B.5.2.1 Tâche 1 : impact des conditions d'acquisition sur les performances des systèmes

- Les coordonnées sont les seuls paramètres considérés dans cette tâche ;
- Les participants ont optimisé leur système sur la base de développement DS2-50, acquise sur une tablette graphique.

Protocole d'évaluation : Les systèmes développés sur DS2-50 ont été testés par l'organisateur sur DS2-382 et DS3-382 pour étudier l'impact des conditions d'acquisition sur les performances des systèmes. Les tests s'effectuent uniquement sur la Session 1.

B.5.2.2 Tâche 2 : impact de la variabilité temporelle sur les performances des systèmes en fonction des paramètres en entrée

- Les participants ont utilisé les deux bases de développement DS2-50 et DS3-50 ; pour DS2-50, trois configurations de paramètres en entrée ont été considérées :

a. Coordonnées seules ;
b. Coordonnées et pression ;
c. Coordonnées, pression et angles d'inclinaison.

– Les participants ont soumis pour chaque base de développement le meilleur système possible en termes de robustesse à la variabilité temporelle ;

Protocole d'évaluation : Les systèmes développés sur DS3-50 ont été testés par l'organisateur sur DS3-382, en ne considérant en entrée du système que les coordonnées (les seuls paramètres capturés par un PDA). Les systèmes développés sur DS2-50 ont été testés par l'organisateur sur DS2-382, en considérant les trois configurations de paramètres. Les tests ont été effectués sur les deux sessions de DS2 et DS3 séparément afin d'étudier l'impact de la variabilité temporelle sur les performances des systèmes.

B.5.2.3 Tâche 3 : impact du contenu d'information dans les signatures sur les performances des systèmes

Protocole d'évaluation : Les systèmes développés sur DS2-50 soumis à la Tâche 1 ont été testés par l'organisateur sur DS2-382 pour chaque catégorie de personnes. Ces catégories ont été générées sur DS2-382 en appliquant une classification hiérarchique sur les 382 valeurs d'Entropie Personnelle. Les tests ont été effectués uniquement sur la Session 1 de la base DS2-382 en ne considérant que les coordonnées en entrée des systèmes.

B.6 Résultats de la Tâche 1

Dans cette tâche, 12 systèmes ont été soumis. Les systèmes ont été développés sur DS2-50, puis testés sur la Session 1 de DS2-382 et DS3-382 (DS2-382 et DS3-382 contiennent les mêmes personnes).

Les Figures B.1 et B.2 montrent les courbes de performances obtenues respectivement sur DS2-382 et DS3-382 avec les deux types d'imitations (bonnes et aléatoires). Les taux d'erreur à l'*EER* sont indiqués dans le Tableau B.2.

Comme on pouvait s'y attendre, les performances sur DS2-382 sont globalement meilleures que sur DS3-382. En considérant les bonnes imitations, les performances sur DS3-382 se dégradent approximativement par un facteur 2 à l'*EER*. En considérant les imitations aléatoires, la dégradation est moins importante.

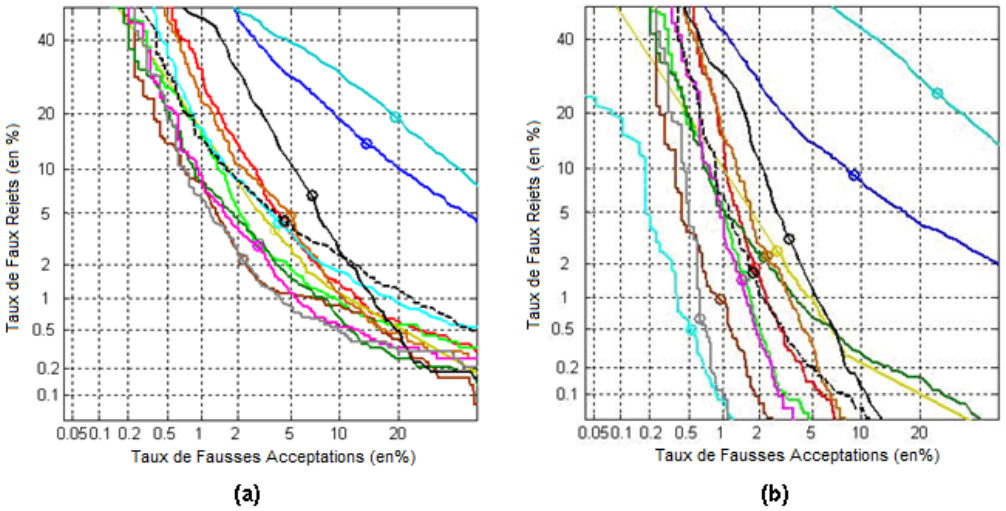

FIGURE B.1: Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382 en considérant (a) les bonnes imitations et (b) les imitations aléatoires.

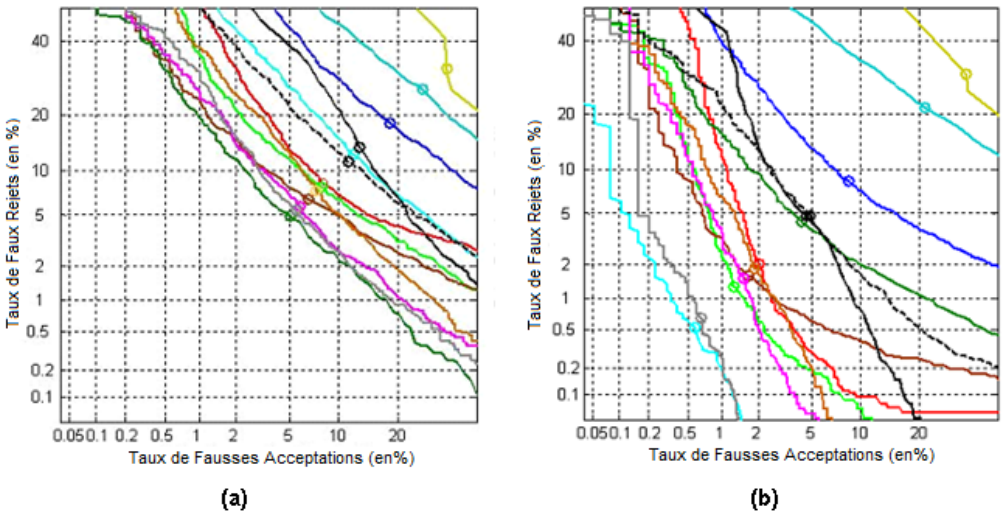

FIGURE B.2: Courbes de performance des systèmes soumis à la Tâche 1 sur DS3-382 en considérant (a) les bonnes imitations et (b) les imitations aléatoires.

Les mauvaises performances obtenues sur DS3-382 sont dues d'une part à la dégradation de la qualité des signatures acquises sur le PDA, comme nous allons le démontrer plus tard dans la Section 7.2 avec notre mesure d'Entropie Personnelle ; et d'autre part, à la bonne qualité des imitations dans la base BioSecure DS3 comme nous l'avons montré avec notre mesure de qualité des imitations dans le paragraphe 6.2.2.3 du Chapitre 6. En effet, les imitations disponibles dans la base BioSecure DS3 [90, 54] ont été acquises suivant un protocole spécifique : l'imposteur visualise sur l'écran tactile du PDA la séquence d'écriture de la signature cible, puis signe directement sur l'image représentant la trajectoire du stylet (voir la description de la base DS3 dans le paragraphe 2.4.4.2 du Chapitre 2).

ID	Tâche 1			
	Test sur DS2-382		Test sur DS3-382	
	Bonnes imitations	Imitations aléatoires	Bonnes imitations	Imitations aléatoires
1	4,4%	1,85%	8,18%	2,05%
2	4,91%	2,33%	7,38%	1,86%
3	13,99%	8,98%	18,32%	8,36%
4	2,88%	1,58%	7,87%	1,29%
5	3,82%	2,67%	31,57%	30,64%
6	2,20%	0,97%	6,58%	1,65%
7	2,98%	2,23%	4,99%	4,32%
8	4,18%	0,51%	12,20%	0,55%
9	2,88%	1,47%	5,77%	1,54%
10	19,23%	24,14%	25,85%	21,34%
11	6,71%	3,31%	13,26%	4,7%
12	2,23%	0,63%	5,47%	0,66%
Ref	4,47%	1,74%	11,27%	4,8%

TABLEAU B.2: Taux d'erreur à l'*EER* sur la Session 1 de DS2-382 et DS3-382 avec les deux types d'imitations, en considérant les coordonnées en entrée des systèmes.

B.7 Résultats de la Tâche 2

B.7.1 Configuration 1 : Coordonnées seulement

Tests sur la base DS2-382 Pour cette tâche, nous avons considéré les 12 systèmes soumis à la Tâche 1 développés sur DS2-50. Dans la Tâche 1, ces systèmes ont été testés sur DS2-382 en considérant uniquement la Session 1. Dans cette tâche, ces systèmes sont testés sur DS2-382 en considérant les 2 sessions séparément.

La Figure B.3 montre les courbes de performances obtenues sur la Session 2 de DS2-382 avec les deux types d'imitations (bonnes et aléatoires). Les taux d'erreur à l'*EER* sont indiqués dans le Tableau B.3.

Pour rappel, dans l'Evaluation 1, les signatures de référence appartiennent à la Session 1, et les tests ont été effectués uniquement sur la Session 1 (absence de variabilité dans le temps). Dans l'Evaluation 2, les signatures de référence appartiennent toujours à la Session 1, mais les tests sont effectués sur la Session 2 (présence de variabilité dans le temps). Ainsi, pour évaluer l'impact de la variabilité temporelle présente dans DS2-382 sur les performances des systèmes, nous comparons les résultats illustrés sur la Figure B.3 et le Tableau B.3 à ceux montrés sur la Figure B.1 et le Tableau B.2 dans la Tâche 1.

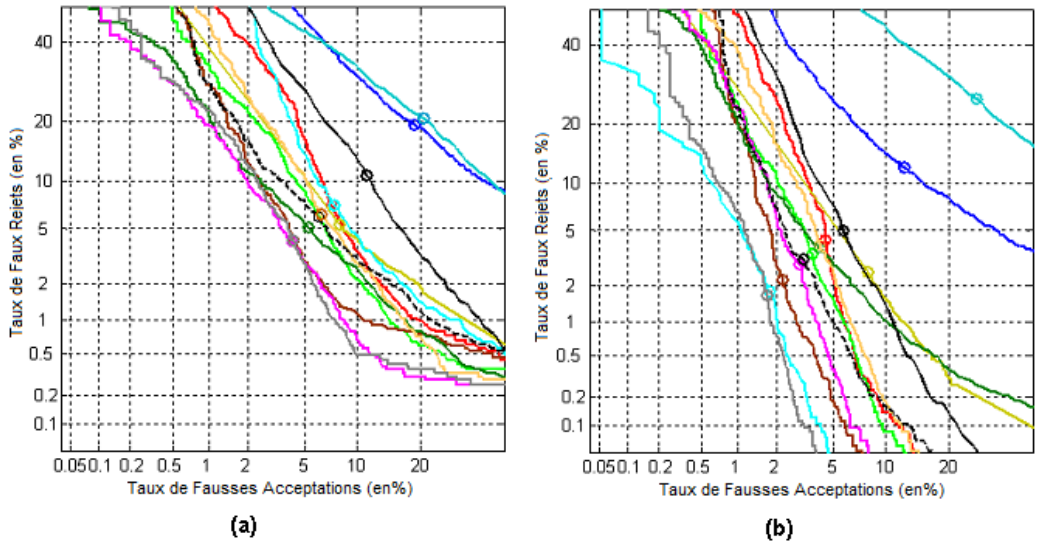

FIGURE B.3: Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires, sur la Session 2 de la base DS2-382 en considérant uniquement les coordonnées en entrée des systèmes.

ID	DS2-382	
	Bonnes imitations	Imitations aléatoires
1	7,15%	4,40%
2	6,20%	4,02%
3	19,03%	12,29%
4	5,99%	3,55%
5	6,61%	5,23%
6	4,21%	2,24%
7	5,13%	3,96%
8	7,26%	1,80%
9	4,08%	2,93%
10	20,47%	25,62%
11	10,92%	5,37%
12	4,18%	1,70%
Ref	5,99%	3,16%

TABLEAU B.3: Taux d'erreur à l'*EER* sur la Session 2 de DS2-382 avec les deux types d'imitations, en considérant les coordonnées en entrée des systèmes.

On observe que pour les deux types d'imitations, les performances se dégradent en présence de la variabilité temporelle.

On note aussi que le classement des systèmes, en se basant sur la valeur du *EER*, ne change pas en présence de la variabilité temporelle.

Tests sur la base DS3-382 Dans cette tâche, 13 systèmes ont été soumis. Les systèmes ont été optimisés sur DS3-50, puis testés sur DS3-382 en considérant les

deux sessions séparément.

Les Figures B.4 et B.5 montrent les courbes de performances obtenues respectivement sur la Session 1 (sans variabilité) et la Session 2 (avec variabilité) de DS3-382 avec les deux types d'imitations (bonnes et aléatoires). Les taux d'erreur à l'*EER* sont indiqués dans le Tableau B.4.

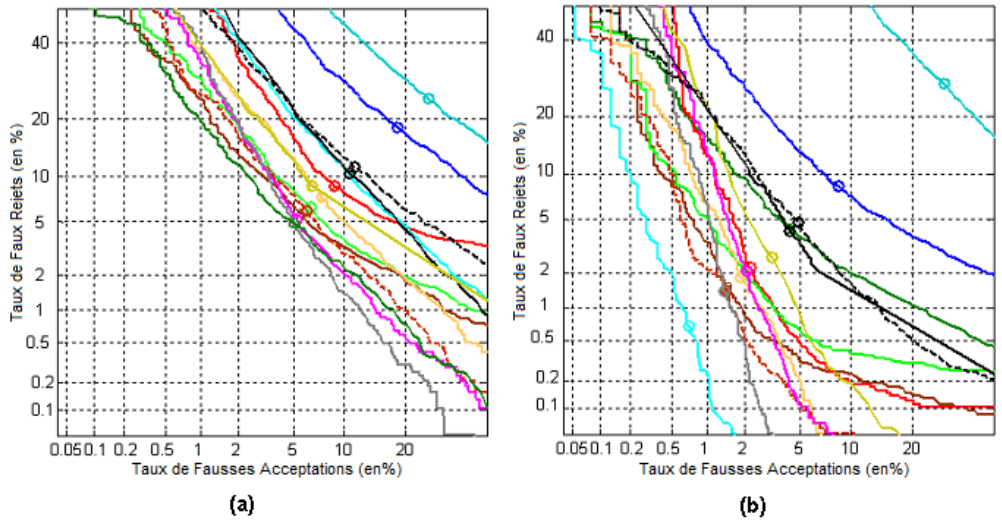

FIGURE B.4: Courbes de performance des systèmes avec (a) les bonnes imitations et (b) les imitations aléatoires, sur la Session 1 de la base DS3-382.

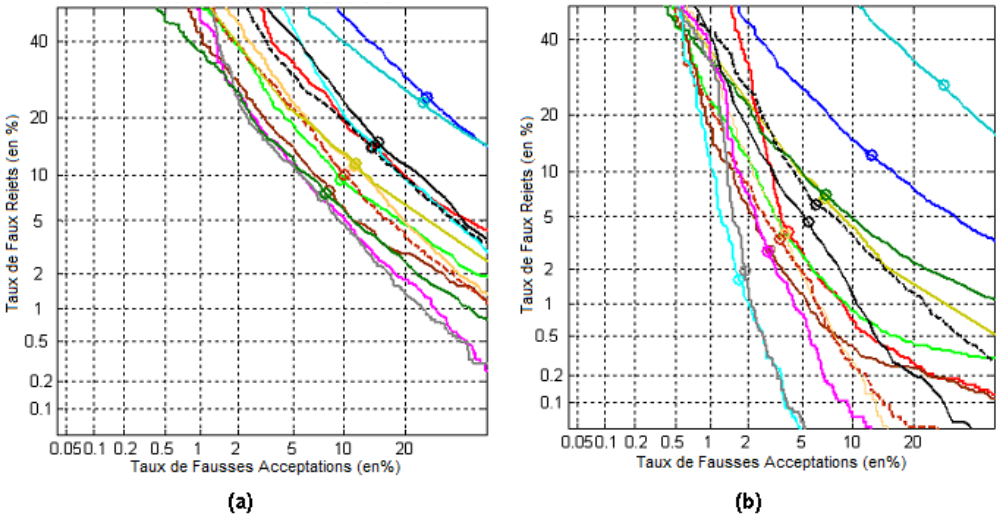

FIGURE B.5: Courbes de performance des systèmes avec (a) les bonnes imitations et (b) les imitations aléatoires, sur la Session 2 de la base DS3-382.

Comme pour DS2-382, on constate que les performances se dégradent en présence de variabilité temporelle pour les deux type d'imitations.

En comparant les systèmes optimisés sur DS2-50 puis testés sur DS3-382 (Tâche 1, Figure B.2) avec ceux optimisés sur DS3-50 puis testés sur DS3-382

ID	DS3-382			
	Test sur Session 1		Test sur Session 2	
	Bonnes imitations	Imitations aléatoires	Bonnes imitations	Imitations aléatoires
1	8,71%	2,22%	14,24%	3,94%
2	7,38%	1,85%	11,25%	3,76%
3	18,32%	8,36%	24,68%	12,40%
4	6,37%	2%	9,43%	3,72%
5	7,55%	4,24%	11,51%	6,78%
6	5,69%	1,5%	8,06%	2,90%
7	4,98%	4,31%	7,69%	7,02%
8	10,40%	0,7%	14,51%	1,67%
9	5,24%	2,09%	7,42%	2,83%
10	24,79%	27,29%	23,52%	26,81%
11	10.49%	2.93%	15%	5.01%
12	4.93%	1.41%	7.42%	1.93%
13	5.98%	1.44%	9.93%	3.48%
Ref	11.27%	4.8%	14.03%	6.06%

TABLEAU B.4: Taux d'erreur à l'*EER* sur les deux Sessions de DS3-382 avec les deux types d'imitations.

(Tâche 2, Figure B.4), on remarque que certains systèmes s'améliorent en termes de performances lorsque la base de test contient des signatures de même nature que celles de la base de développement.

B.7.2 Configuration 2 : Coordonnées et pression

Tests sur la base DS2-382 seulement Pour cette tâche, 11 systèmes ont été soumis. Ils ont été optimisés sur DS2-50 en considérant les coordonnées et la pression comme paramètres d'entrée. Ces systèmes sont testés sur DS2-382 en considérant les deux sessions séparément.

Les Figures B.6 et B.7 montrent les courbes de performances obtenues respectivement sur la Session 1 (sans variabilité) et la Session 2 (avec variabilité) de DS2-382 avec les deux types d'imitations (bonnes et aléatoires). Les taux d'erreur à l'*EER* sont indiqués dans le Tableau B.5.

Dans le cas d'absence de variabilité temporelle, en comparant la Figure B.1 où seules les coordonnées sont pris en compte à la Figure B.6 où la pression a été ajoutée, on remarque que les performances des systèmes s'améliorent lorsque l'on considère en plus des coordonnées le paramètre pression.

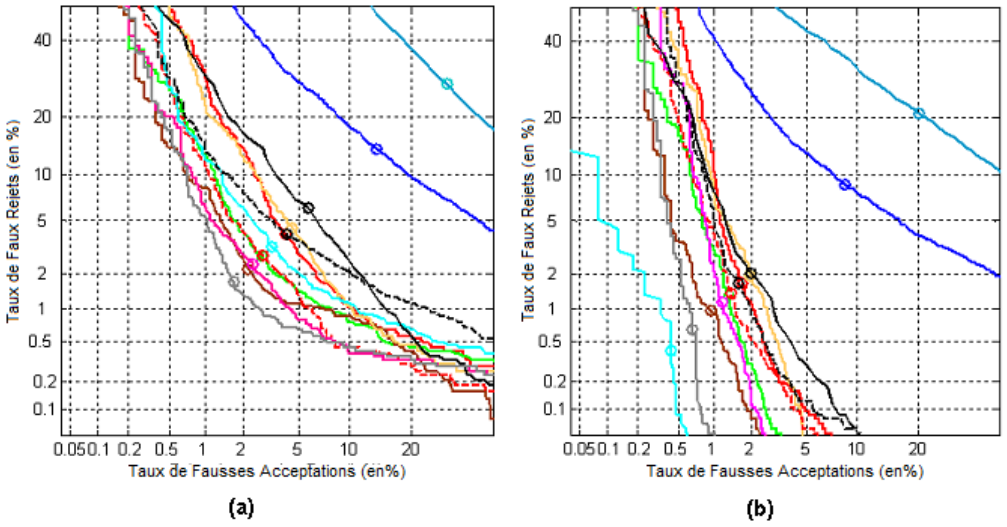

FIGURE B.6: Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires sur la Session 1 de DS2-382, en considérant les coordonnées et la pression en entrée des systèmes.

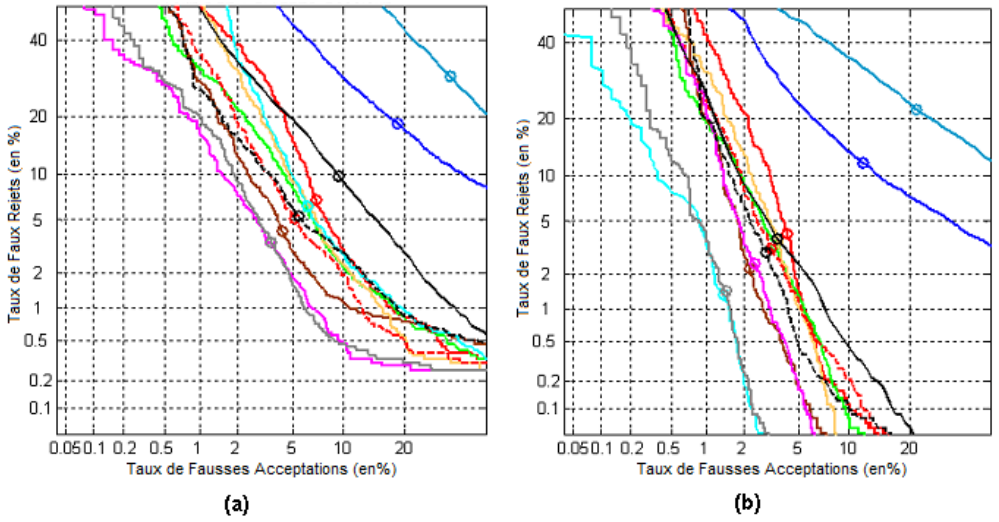

FIGURE B.7: Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires sur la Session 2 de DS2-382, en considérant les coordonnées et la pression en entrée des systèmes.

Cependant, d'après les Figures B.6 et B.7, la pression n'améliore pas la robustesse des systèmes à la variabilité temporelle.

ID	DS2-382			
	Test sur Session 1		Test sur Session 2	
	Bonnes imitations	Imitations aléatoires	Bonnes imitations	Imitations aléatoires
1	4,03%	1,70%	6,88%	4,16%
2	4,5%	1,96%	6,28%	3,61%
3	13,69%	8,62%	18,54%	11,81%
4	2,76%	1,33%	6,07%	3,42%
5	2,19%	0,97%	4,21%	2,23%
6	3,26%	0,42%	6,21%	1,37%
7	2,38%	1,17%	3,48%	2,46%
8	27,76%	20,51%	30,13%	21,61%
9	5,9%	2,02%	9,52%	3,65%
10	1,71%	0,65%	3,49%	1,46%
11	2,84%	1,38%	5,1%	3,19%
12	4,07%	1,65%	5,32%	2,96%

TABLEAU B.5: Taux d'erreur à l'*EER* sur les deux Sessions de DS2-382 avec les deux types d'imitations, en considérant les coordonnées et la pression en entrée des systèmes.

B.7.3 Configuration 3 : Coordonnées, pression et angles d'inclinaison

Tests sur la base DS2-382 seulement Dans cette tâche, 6 systèmes ont été soumis. Ils ont été développés sur DS2-50 en considérant les cinq fonctions temporelles acquises sur la tablette graphique. Ces systèmes sont testés sur DS2-382 en considérant les deux sessions séparément.

Les Figures B.8 et B.9 montrent les courbes de performances obtenues respectivement sur la Session 1 (sans variabilité) et la Session 2 (avec variabilité) de DS2-382 avec les deux types d'imitations (bonnes et aléatoires). Les taux d'erreur à l'*EER* sont indiqués dans le Tableau B.6.

Dans le cas d'absence de variabilité temporelle, en comparant la Figure B.6 où les coordonnées et la pression sont pris en compte à la Figure B.8 où les angles d'inclinaison ont été ajoutés, on remarque que les performances des systèmes ne s'améliorent pas lorsque l'on considère, en plus des coordonnées et la pression, les angles d'inclinaison.

De plus, d'après les Figures B.8 et B.9, les angles d'inclinaison n'améliorent pas la robustesse des systèmes à la variabilité temporelle.

En conclusion, la pression améliore les performances dans le cadre d'absence de variabilité temporelle par rapport aux seules fonctions coordonnées. Cependant, ceci cesse d'être le cas dans un contexte de variabilité temporelle à "long-terme". De plus, les angles d'inclinaison sont de mauvais paramètres en présence ou pas de variabilité temporelle.

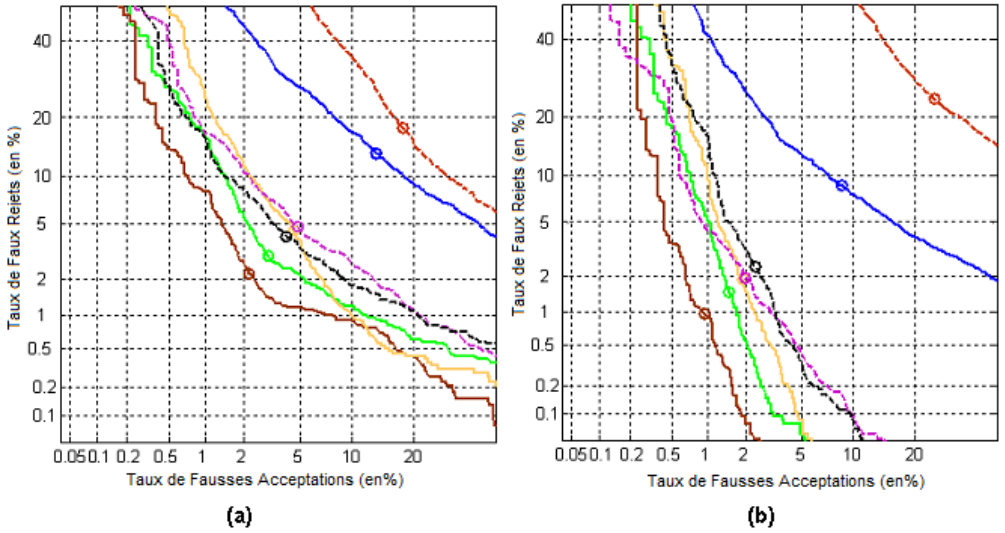

FIGURE B.8: Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires sur la Session 1 de DS2-382, en considérant les coordonnées, la pression et les angles d'inclinaison en entrée des systèmes.

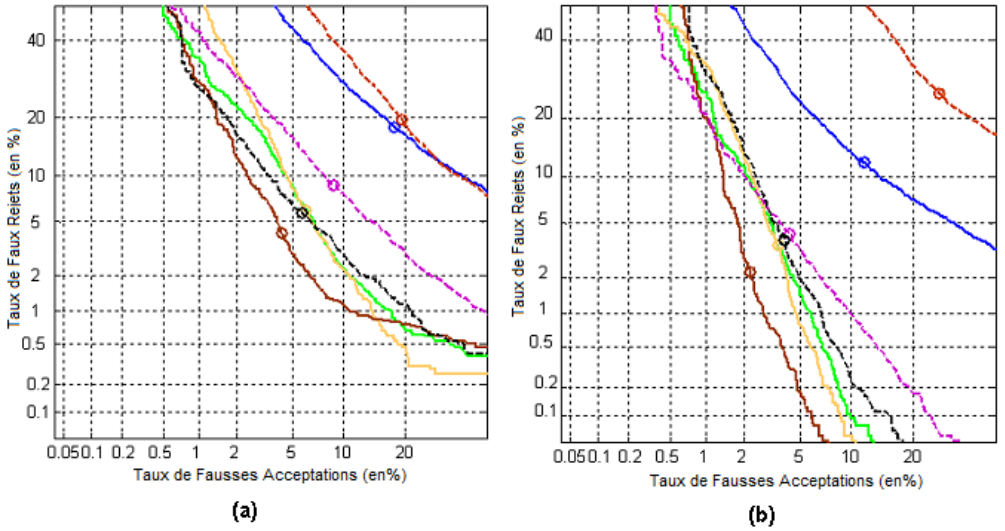

FIGURE B.9: Courbes de performance des systèmes soumis à la Tâche 2 avec (a) les bonnes imitations et (b) les imitations aléatoires sur la Session 2 de DS2-382, en considérant les coordonnées, la pression et les angles d'inclinaison en entrée des systèmes.

ID	DS2-382			
	Test sur Session 1		Test sur Session 2	
	Bonnes imitations	Imitations aléatoires	Bonnes imitations	Imitations aléatoires
2	4,52%	1,91%	5,99%	3,53%
3	13,41%	8,63%	17,91%	11,77%
4	3,02%	1,49%	6,02%	3,52%
6	2,19%	0,97%	4,21%	2,23%
13	17,94%	24,06%	19,34%	25,61%
14	4,82%	1,98%	8,73%	4,24%
Ref	4,07%	2,39%	5,72%	3,87%

TABLEAU B.6: Taux d'erreur à l'*EER* sur les deux Sessions de DS2-382 avec les deux types d'imitations, en considérant les coordonnées, la pression et les angles d'inclinaison en entrée des systèmes.

B.8 Résultats de la Tâche 3

Pour cette évaluation, nous avons considéré les deux catégories extrêmes seulement. La première correspond à la catégorie "Haute Entropie". Elle contient les signatures de 60 personnes. La deuxième correspond à la catégorie "Basse Entropie". Elle contient les signatures de 161 personnes.

Les Figures B.10 et B.11 montrent les courbes de performance obtenues sur la Session 1 de DS2-382 pour les deux catégories extrêmes d'Entropie, avec les deux types d'imitations (bonnes et aléatoires). Les taux d'erreur à l'*EER* sont indiqués dans le Tableau B.7.

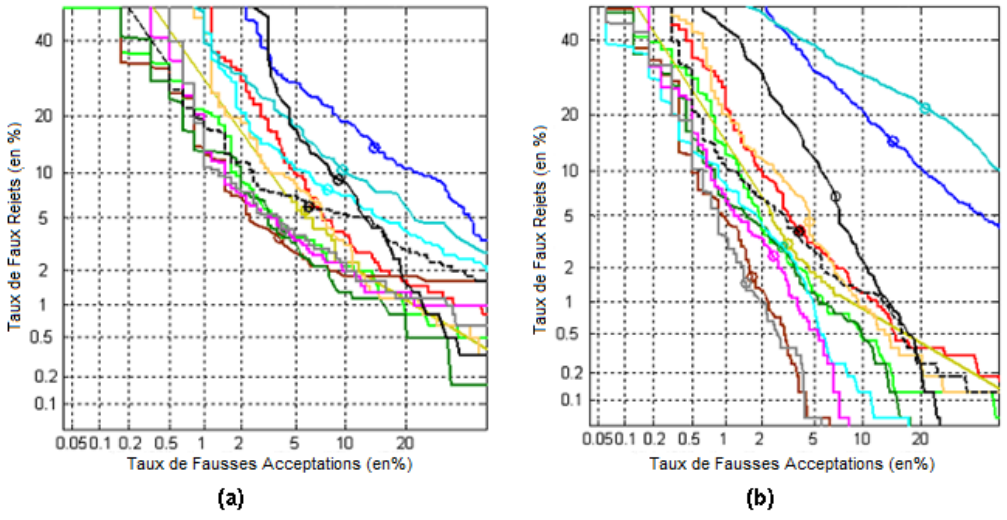

FIGURE B.10: ourbes de performance des systèmes soumis à la Tâche 1 sur DS2-382, en considérant les bonnes imitations, pour les 2 catégories extrêmes d'Entropie : (a) Haute et (b) Basse Entropie.

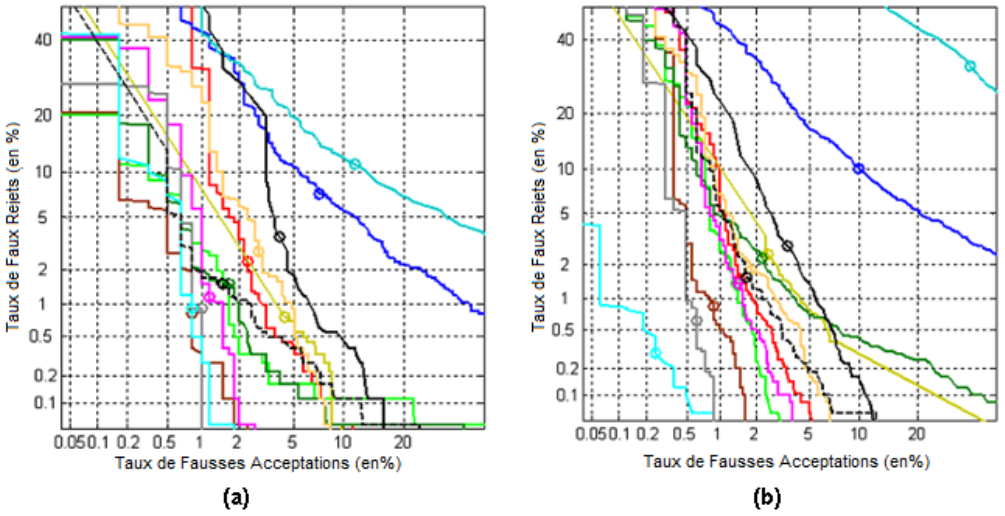

FIGURE B.11: Courbes de performance des systèmes soumis à la Tâche 1 sur DS2-382, en considérant les imitations aléatoires, pour les 2 catégories extrêmes d'Entropie : (a) Haute et (b) Basse Entropie.

ID	DS2-382			
	Haute Entropie		Basse Entropie	
	Bonnes imitations	Imitations aléatoires	Bonnes imitations	Imitations aléatoires
1	6,5%	2,33%	3,94%	1,50%
2	6,58%	2,83%	4,57%	1,80%
3	14,00%	7,22%	14,5%	9,98%
4	4,08%	1,52%	2,92%	1,38%
5	5,67%	2,55%	3,14%	2,47%
6	3,75%	0,83%	1,68%	0,87%
7	4,00%	1,61%	2,89%	2,27%
8	7,83%	0,8%	2,95%	0,27%
9	4,17%	1,19%	2,48%	1,42%
10	9,92%	11,27%	21,18%	32,43%
11	9,00%	3,83%	6,83%	3,14%
12	4,17%	0,91%	1,49%	0,62%
Ref	6,00%	1,52%	3,81%	1,62%

TABLEAU B.7: Taux d'erreur à l'*EER* sur la Session 1 de DS2-382 pour les deux catégories extrêmes de personnes avec les deux types d'imitations.

Hormis les deux systèmes qui ont montré des problèmes lors de l'évaluation BSEC'2009 (Système 2 et Système 10), pour tous les autres systèmes soumis, les résultats montrent qu'il existe une différence significative des performances entre les deux catégories extrêmes, pour les deux types d'imitations (bonnes et aléatoires) : les meilleures performances sont obtenues sur les personnes appartenant à la catégorie "Basse Entropie", dont les signatures sont les plus longues, les plus complexes et les plus stables. A l'opposé, les performances se dégradent sur les personnes appartenant à la catégorie "Haute Entropie", dont les signatures sont les plus courtes, les moins complexes et les plus variables. Nous avons toutefois

noté que certains systèmes sont plus robustes que d'autres à cette dégradation de qualité. De plus, le classement des systèmes change selon la catégorie d'Entropie Personnelle considérée pour le test.

Ces résultats viennent confirmer sur de nombreux systèmes de vérification et de grandes bases de signatures les résultats que nous avons obtenus avec différentes approches de vérification au préalable [49, 43, 42]. Ainsi, la capacité de notre mesure d'Entropie Personnelle à catégoriser les utilisateurs apparaît clairement.

Annexe C

Liste des publications

Articles de revues

S. Garcia-Salicetti, N. Houmani, B. Dorizzi, "A Novel Criterion for Writer Enrolment based on a Time-Normalized Signature Sample Entropy Measure", EURASIP Journal on Advances in Signal Processing, 2009.

Chapitres de livres

S. Garcia-Salicetti, N. Houmani, B. Ly-Van, B. Dorizzi, F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, C. Vielhauer and T. Scheidat, "Online Handwritten Signature Verification", D. Petrovska-Delacretaz and G. Chollet and B. Dorizzi (Eds.), Guide to Biometric Reference Systems and Performance Evaluation, Springer-Verlag, London, 2008 (ISBN 978-1-84800-291-3).

S. Garcia-Salicetti, N. Houmani, "Digitizing Tablet", Z. Li Stan (Ed.), Encyclopedia of Biometrics, Springer-Verlag, Germany, 2009.

Articles de conférences

S. Garcia-Salicetti, N. Houmani, B. Dorizzi, "A Client-entropy Measure for On-line Signatures", Proc. of IEEE Biometrics Symposium (BSYM'08), Tampa, USA, September 2008.

N. Houmani, S. Garcia-Salicetti, B. Dorizzi, "A Novel Personal Entropy Measure confronted with Online Signature Verification Systems' Performance", Proc. of IEEE Second International Conference on Biometrics : Theory, Applications and Systems (BTAS'08), Washington, September 2008.

B. Dorizzi, R. Cappelli, M. Ferrara, D. Maio, D. Maltoni, N. Houmani, S. Garcia-Salicetti and A. Mayoue, “Fingerprint and On-Line Signature Verification Competitions at ICB 2009”, In Proceedings of International Conference on Biometrics (ICB’09), LNCS Vol. 5558, Springer, p.725-732, June 2009.

N. Houmani, S. Garcia-Salicetti, B. Dorizzi, “On assessing the Robustness of Pen Coordinates, Pen Pressure and Pen inclination to Time Variability with Personal Entropy”, Proc. of IEEE Third International Conference on Biometrics : Theory, Applications and Systems (BTAS’09), Washington, September 2009.

J. Montalvão, N. Houmani, B. Dorizzi, “Comparing GMM and Parzen in Automatic Signature Verification - A Step Backward or Forward ?”, Proc. of XVIII Congresso Brasileiro de Automatica (CBA’10), September 2010.

N. Houmani, S. Garcia-Salicetti, B. Dorizzi, Mounim El-Yacoubi, “On-line Signature Verification on a Mobile Platform”, International Workshop on Mobile Security, Santa Clara, USA, October 2010 .