

Vérification d'Identité par Ecriture et Parole Combinées

Jean Hennebert¹ – Andreas Humm¹ – Rolf Ingold¹

Groupe DIVA, Département d'Informatique
Boulevard de Pérolles 90, 1700 Fribourg, Suisse

jean.hennebert, andreas.humm, rolf.ingold@unifr.ch

Résumé : Nous rapportons les premiers développements d'un système de vérification d'identité par utilisation combinée de l'écriture et de la parole. La nouveauté de notre approche réside dans l'enregistrement simultané de ces deux modalités en demandant à l'utilisateur d'énoncer ce qu'il est en train d'écrire. Nous présentons et analysons deux scénarii : la signature lue où l'utilisateur énonce le contenu de sa signature et l'écriture lue. Nous décrivons le système d'acquisition, l'enregistrement d'une base de données d'évaluation, les résultats d'une enquête d'acceptabilité, le système de vérification à base de multi-gaussiennes et les résultats de ce dernier obtenus pour le scénario signature.

Mots-clés : Vérification de signature, d'écriture, de la parole, biométrie multimodale.

1 Introduction

La parole et l'écriture sont deux modalités majeures utilisées quotidiennement et naturellement par la plupart d'entre nous. Nous proposons un système de vérification automatique d'identité utilisant de façon combinées l'écriture en ligne et la parole. La nouveauté de notre approche réside dans l'acquisition des signaux qui est ici simultanée. En d'autres termes, l'utilisateur est prié d'énoncer de façon aussi synchrone que possible ce qu'il est en train d'écrire. Nous avons nommé notre approche *Modalités Ecriture-Parole Combinées (CHASM¹)* pour mettre en évidence le fait que ces deux modalités sont enregistrées en même temps.

Deux scénarii de vérification d'identité ont été définis et sont à la base de nos développements. Pour le premier scénario, nous évaluons la faisabilité d'une *signature CHASM*, où l'utilisateur énonce le contenu de sa signature. Pour le deuxième scénario, l'utilisateur lit et écrit un morceau de texte - *écriture CHASM*. Nous soulignons que les enregistrements CHASM pourraient également être utilisés pour effectuer de la reconnaissance automatique de contenu. Ceci est néanmoins hors du contexte de cet article où nous utilisons les données CHASM pour des applications biométriques.

Cet article est organisé de la façon suivante. Dans cette section d'introduction, nous présentons les motivations principales de CHASM et nous faisons référence à certains travaux connectés à notre approche. Dans la section 2, nous présentons le système d'acquisition, l'enregistrement d'une base de données d'évaluation ainsi que les résultats d'une enquête d'acceptabilité menée lors de l'enregistrement de la base de

données. Dans la section 3, nous décrivons le système de vérification basé sur des multi-gaussiennes (GMMs²). Dans la section 4, nous présentons les résultats obtenus par le système GMMs sur le scénario *signature CHASM*. Finalement, des conclusions et nos travaux futurs sont présentés.

1.1 Motivations

De nombreux systèmes biométriques utilisant la signature, l'écriture ou la parole de façon isolée ont été étudiés [LEC 94, PLA 94, REY 02, AL 04]. Néanmoins, nous observons que ces modalités restent faiblement utilisées dans des applications biométriques commerciales. Nous pouvons avancer trois raisons majeures à ceci : (1) la grande variabilité temporelle intra-utilisateur implique souvent l'acquisition de sessions d'enrôlement espacées dans le temps afin d'atteindre des performances suffisantes [LYV 03], (2) il est possible de commettre des impostures en s'entraînant aux imitations (essentiellement pour la signature et l'écriture [LEE 96, VIE 06]), (3) les performances se dégradent lorsque les conditions d'enrôlement et de vérification varient, comme par exemple senseurs ou environnement différents. Si les facteurs (2) et (3) peuvent être minimisés en supervisant les acquisitions et en contrôlant l'environnement, le point (1) reste un problème essentiel.

Les approches *signature CHASM* et *écriture CHASM* que nous proposons ici visent à réduire les problèmes énoncés ci-dessus tout en restant acceptables pour l'utilisateur final. Tout d'abord, nous visons des performances meilleures grâce à la bi-modalité. Nous pourrions vraisemblablement réduire la durée ou le nombre de sessions d'enrôlement pour des performances équivalentes. Ensuite, la robustesse contre les imitations entraînées sera potentiellement meilleure, étant donné que l'imposteur devra faire face à une charge cognitive supplémentaire en tentant d'imiter à la fois l'écriture et la voix d'un individu. Finalement, d'un point de vue pratique, les senseurs requis pour notre approche existent et sont peu coûteux.

1.2 Travaux connexes

Dans [FUE 02], un système de vérification de signature en ligne et un système de vérification du locuteur, tous deux basés sur l'utilisation de modèles de Markov cachés (HMMs), sont fusionnés au niveau des scores de vraisemblance. Les résultats obtenus par la fusion sont meilleurs que pour les systèmes évalués de façon isolée. Pour réali-

¹Combined Handwriting And Speech Modalities

²Gaussian Mixture Modelling

ser ces tests, des utilisateurs fictifs sont artificiellement créés en associant aléatoirement des signatures et des échantillons de parole pris dans deux bases de données indépendantes. Dans [LYV 03], des conclusions similaires sont rapportées sur des données de signature et de parole provenant du même utilisateur enregistré dans le cadre de la base de données BIOMET [GAR 03]. Dans [KRA 05], un système sur tablet-PC utilisant la parole et la signature en ligne est proposé pour sécuriser l'accès aux données médicales. L'argument principal est d'utiliser les modalités offertes de base sur les tablet-PCs dorés et déjà adoptés par de nombreux membres du corps médical.

La principale différence entre ces travaux et notre approche CHASM réside dans la phase d'acquisition. Dans notre cas, les signaux d'écriture et de parole sont enregistrés de façon simultanée, en demandant à l'utilisateur d'énoncer ce qu'il est en train d'écrire. Notre approche a l'avantage clair de réduire le temps d'enrôlement et d'accès. Également, la relative synchronisation des deux signaux ouvre éventuellement la porte à des stratégies de fusion plus robustes intervenant à des stades plus en amont de la chaîne de traitement (fusion de caractéristiques, modélisation jointe, etc).

2 Acquisition de données CHASM

2.1 Système d'acquisition

Nous avons élaboré un système d'acquisition utilisant une tablette graphique WACOM Intuos2 et un microphone standard. Le signal d'écriture est échantillonné à 100 Hz et comprend les coordonnées (x, y) , la pression et les angles d'élévation et d'azimut du stylo. Le signal de parole est échantillonné à 16 kHz et codé linéairement sur 16 bits. Les temps d'acquisition des échantillons d'écriture et de parole sont également enregistrés afin de pouvoir synchroniser les deux flux de données. La figure 1 montre un exemple d'une signature CHASM. Les parties grisées correspondent à des moments où le stylo est hors de portée de la tablette.

2.2 Base de données

Des données CHASM ont été enregistrées dans le cadre du projet MyIDEa visant la collecte d'une base de données biométrique largement multimodale [AL 05, HEN 05]. La base de données MyIDEa contient d'autres modalités telles que les empreintes digitales, des séquences vidéo du visage, des images de la paume de la main, etc. La base de données contient 70 utilisateurs qui ont été enregistrés lors de trois sessions espacées dans le temps. Le jeu de données MyIDEa incluant les données CHASM utilisées dans cet article porte la référence MYIDEA-CHASM-SET1. Les données CHASM ont été enregistrées suivant les deux scénarii *signature CHASM* et *écriture CHASM* mentionnés auparavant.

Signature CHASM. Six signatures sont enregistrées par session. Il est également demandé à chaque sujet d'imiter six fois la signature d'un autre sujet après avoir inspecté une version statique de la signature. La voix n'est par contre pas imitée mais le contenu vocal de la signature à imiter est reproduit. Cette procédure donne un total de 18 accès *vrais* et de 18 *impostures entraînées*.

Ecriture CHASM Pour chaque session, le texte repro-

duit par l'utilisateur est composé d'une phrase fixe contenant toutes les lettres de l'alphabet et d'un morceau de texte de 50 à 100 mots choisi aléatoirement dans un corpus. La phrase fixe est utilisée pour évaluer les systèmes dépendants du texte tandis que la partie aléatoire est utilisée pour évaluer les systèmes indépendants du texte. De façon similaire à l'acquisition de signature, le sujet imite l'écriture d'un autre utilisateur après inspection d'une version statique de l'écriture, sans tenter d'imiter la partie parole. Cette procédure donne un total de 3 accès *vrais* et de 3 *impostures entraînées*.

Nous avons noté que tous les utilisateurs, sans exception, sont parvenus à faire les acquisitions. Devoir parler et écrire en même temps n'a empêché aucune acquisition de se produire. Pour la signature, la majorité des utilisateurs ont lu effectivement le contenu de leur signature, synchronisant les symboles écrits avec des syllabes. Les symboles "ornementaux" qui commencent ou terminent la plupart des signatures n'étaient spontanément pas lus par les utilisateurs. Il est à remarquer que quelques utilisateurs présentaient des signatures composées essentiellement de symboles non lisibles. Il était alors demandé à ces utilisateurs de simplement dire leur nom en même temps que signer. Pour l'écriture CHASM, nous avons remarqué que la voix est légèrement désynchronisée de l'écriture. Des points de synchronisation apparaissent en moyenne au moment des syllabes et des fins de mots. Certains utilisateurs prononcent spontanément la ponctuation, d'autres ne la prononce pas.

2.3 Enquête d'acceptabilité

Nous avons demandé aux 70 sujets enregistrés dans la base de données de répondre à un ensemble de questions relatives à la procédure CHASM :

1. Trouvez-vous simple/difficile d'écrire sur une tablette ?
2. Pensez-vous avoir écrit plus vite, à la même vitesse ou plus lentement que d'habitude ?
3. Trouvez-vous simple/difficile de parler et signer en même temps ?
4. Trouvez-vous simple/difficile de parler et écrire en même temps ?
5. Combien de lignes de texte accepteriez-vous d'écrire pour vous identifier dans le cadre d'un accès à votre compte bancaire ?
6. Pensez-vous que le fait de parler et écrire en même temps a affecté vos capacités d'imiter l'écriture de quelqu'un d'autre ?

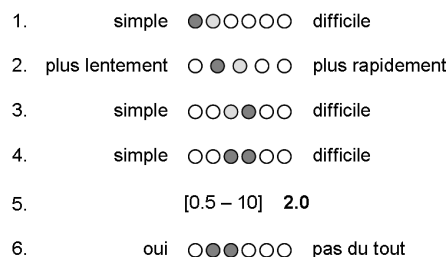


FIG. 2 – Résultats de l'enquête d'acceptabilité.

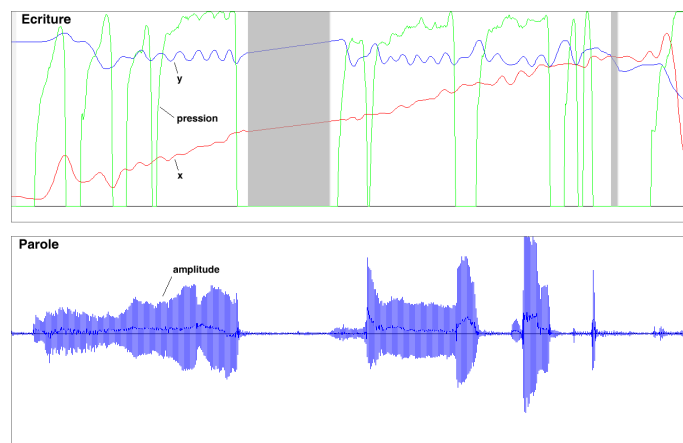


FIG. 1 – Visualisation synchronisée des signaux d’écriture (haut) et de parole (bas) dans le cas d’une signature. Les angles d’azimut et d’élévation du stylo ne sont pas montrés par soucis de clarté.

Bien que limitée à environ 70 utilisateurs, cette enquête a le mérite de nous apporter des renseignements au niveau de la perception ressentie face à une acquisition CHASM. La figure 2 donne une vue schématique des réponses données par les sujets. Ecrire sur une tablette est jugé simple par la plupart des utilisateurs. Ecrire et parler en même temps est jugé de difficulté moyenne. Cette qualification est probablement due au niveau de concentration requis plus élevé que pour une écriture simple. Tous les utilisateurs ont pu signer et dire le contenu de leur signature, néanmoins, l’acquisition de la signature lue est jugée plus difficile que l’écriture lue. Une interprétation peut être donnée en considérant le fait qu’une signature contient souvent des ornements qui ne peuvent être lus. La vitesse d’écriture est jugée plus lente que pour une écriture simple et les utilisateurs accepteraient d’écrire jusqu’à deux lignes de texte pour être identifiés automatiquement. Les utilisateurs ressentent que le fait de parler en même temps a affecté leur capacité à imiter la signature et l’écriture d’un autre. D’un point de vue objectif, nous avons constaté que tous les utilisateurs ont pu réaliser les acquisitions CHASM. Les réponses de l’enquête expriment de façon subjective une certaine difficulté associée aux acquisitions, mais n’indiquent pas un rejet de la méthodologie. Nous en concluons que notre approche est acceptable du point de vue production et enregistrement des données CHASM.

3 Description du système de vérification

La figure 3 illustre notre système de vérification. Dans l’optique de mettre au point un système de référence simple, nous avons choisi d’utiliser des GMMs standards pour modéliser de façon indépendantes les deux flux de signaux. La fusion s’effectue au niveau des scores de vraisemblance. Cette approche permet également de mesurer les performances des sous-systèmes parole seule (1), écriture seule (2) et fusion (3).

L’architecture de ce système est essentiellement non spécifique et s’applique sans modifications majeures aux deux scénarii *signature CHASM* et *écriture CHASM*. Néanmoins, dans cet article, nous rapportons uniquement les résultats de

l’évaluation du scénario *signature CHASM*.

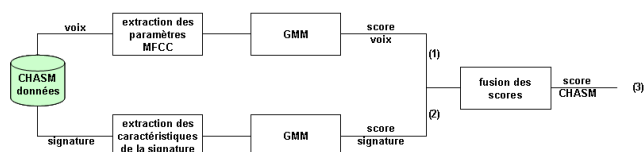


FIG. 3 – Système de base pour la vérification utilisant l’approche CHASM.

3.1 Extraction des paramètres d’écriture

Pour chaque point du signal d’écriture en ligne, nous extrayons 25 paramètres, d’une façon similaire à ce qui est décrit dans [LYV 04] :

- la vitesse et l’accélération absolue, la vitesse et l’accélération dans les directions x et y , l’accélération tangentielle
- l’angle α du vecteur de vitesse absolue, son cosinus et sinus, la dérivée de α et son cosinus et sinus
- la pression et la dérivée de la pression
- les angles d’azimut et d’élévation et leurs dérivées
- le rayon de courbure
- les coordonnées normalisées $(x(n) - x_g, y(n) - y_g)$ relativement au centre de gravité (x_g, y_g) (pour l’extraction de caractéristiques de signatures uniquement)
- le rapport longueur à largeur d’une fenêtre de 5 et 7 points centrée sur le point courant, ainsi que le rapport vitesse minimale à vitesse maximale calculé sur une fenêtre de 5 points autour du point courant

Ces paramètres sont ensuite normalisés par la moyenne et l’écart type (normalisation z -norm) estimés pour chaque utilisateur.

3.2 Extraction de paramètres de la parole

Nous utilisons des paramètres Mel cepstraux (MFCC) [RAB 93] calculés sur des fenêtres de 25.6 ms décalées de 10 ms. Pour chaque fenêtre, 12 paramètres MFCC sont extraits ainsi que l’énergie du signal dans la fenêtre. Un module de détection de la parole basé sur un

modèle bi-gaussien est utilisé pour enlever les parties de silence du signal. Les MFCC sont ensuite normalisés par la moyenne et l'écart type estimés pour chaque utilisateur sur la partie du signal après suppression du silence.

3.3 Modélisation par multi-gaussiennes

Des multi-gaussiennes (GMMs) sont utilisées pour estimer les vraisemblances des vecteurs de caractéristiques. La modélisation par GMMs n'est sans doute pas la méthode la plus appropriée pour modéliser les signaux dépendants du texte, comme par exemple la signature ou les phrases fixes de nos scénarii. Néanmoins, les GMMs sont souvent considérés comme systèmes de référence pour la modélisation du locuteur et ont donné des résultats comparables aux HMMs pour des tâches de vérification de signature [RIC 03]. Les GMMs sont également des outils de modélisation simples à utiliser et très flexibles, capables de modéliser des densités de probabilités relativement complexes.

Les GMMs estiment la densité de probabilité $p(x_n|M_{client})$, aussi nommée la *vraisemblance*, d'un vecteur à D dimensions par une somme pondérée de densités gaussiennes multivariées :

$$p(x_n|M_{client}) = \sum_{i=1}^I w_i \mathcal{N}(x_n, \mu_i, \Sigma_i) \quad (1)$$

où I est le nombre de gaussiennes et w_i est le poids de la gaussienne i . Les densités gaussiennes $\mathcal{N}$ sont paramétrées par un vecteur de moyenne μ_i ($D \times 1$) et une matrice de covariance Σ_i ($D \times D$). Les poids w_i satisfont la contrainte $\sum_{i=1}^I w_i = 1$.

Nous faisons ici l'hypothèse que les coefficients des vecteurs de paramètres sont décorrélés afin de pouvoir utiliser une matrice de covariance diagonale. En faisant l'hypothèse supplémentaire d'indépendance temporelle des observations, le score global de vraisemblance du modèle du client pour la séquence de vecteur de paramètres $X = \{x_1, x_2, \dots, x_N\}$ est calculé comme suit :

$$S_c = p(X|M_{client}) = \prod_{n=1}^N p(x_n|M_{client}). \quad (2)$$

De façon similaire, nous estimons également la vraisemblance S_w que X n'appartienne pas au client en utilisant un modèle du monde M_{world} . Dans notre cas, ce modèle du monde est unique et est entraîné sur un ensemble de données provenant d'autres utilisateurs du système. En fonction des hypothèses faites précédemment, la décision optimale d'accepter ou de rejeter l'identité prétendue de l'utilisateur est prise en comparant le rapport des scores de vraisemblance du client et du monde avec un seuil global T . Ce rapport est souvent calculé dans le domaine logarithmique :

$$R_c = \log(S_c) - \log(S_w). \quad (3)$$

L'entraînement des modèles du client et du monde se fait via l'algorithme *Expectation-Maximisation* (EM) qui raffine de façon itérative les différents paramètres du modèle afin de maximiser sa vraisemblance [DEM 77]. Les modèles du

client et du monde sont entraînés indépendamment en appliquant la procédure EM jusqu'à la convergence des paramètres, typiquement après quelques itérations. Nous utilisons une procédure d'éclatement binaire des gaussiennes afin d'augmenter le nombre de gaussiennes jusqu'à une valeur prédéfinie. Cette méthode permet d'obtenir des GMMs avec un nombre élevé de gaussiennes tout en limitant les risques de tomber dans des maxima locaux lors de la procédure EM. Le modèle du monde est entraîné en utilisant tous les accès clients disponibles dans la base de données. Les accès imposteurs sont exclus pour l'entraînement de ce modèle du monde.

3.4 Fusion des scores

Les deux modalités sont fusionnées par une simple règle d'addition des scores de vraisemblance $R_{c,CHASM} = R_{c,parole} + R_{c,écriture}$. Cette procédure est raisonnable si nous faisons d'une part l'hypothèse d'estimateurs de vraisemblance non-biaisés et d'autre part l'hypothèse d'indépendance des signaux de parole et d'écriture. Cette dernière hypothèse n'est clairement pas vérifiée dans notre cas étant donné qu'il est spécifiquement demandé aux utilisateurs de synchroniser la parole avec l'écriture. Des approches plus élaborées pour réaliser la fusion pourraient être envisagées, comme par exemple une fusion pondérée des scores ou encore l'utilisation de classificateurs entraînés à fusionner [JAI 05]. Néanmoins, de telles méthodes nécessitent une estimation des paramètres sur un set de données indépendant, non disponibles actuellement.

Nous rapportons donc les résultats de la fusion en utilisant la simple addition de score décrite ci-dessus. Nous rapportons également le résultat d'une normalisation *z-norm* des scores effectuée avant la sommation. La *z-norm* est ici appliquée globalement d'une part sur les scores paroles et d'autre part sur les scores écritures. Étant donné que la moyenne et l'écart-type de la *z-norm* sont ici estimés à posteriori sur le même set de données, les résultats seront optimistes par rapport à une estimation à priori.

4 Evaluation du système

Nous rapportons dans cette section les résultats de l'évaluation du système décrit en section 3 pour le scénario *signature CHASM*. Des protocoles d'évaluation ont été définis pour les scénarii *signature CHASM* et *écriture CHASM* dans [HEN 05]. Ces protocoles, tout en adhérant à des utilisations réalistes du système, visent à mettre en évidence les impacts des impostures entraînées et l'influence de la variabilité temporelle.

4.1 Protocoles d'évaluation

Pour le scénario *signature CHASM*, les protocoles suivants ont été définis :

Sans variabilité temporelle. Le modèle du client est construit à partir de trois signatures de la première session. Les tests d'accès clients sont effectués sur les trois signatures restantes de la première session. La même procédure est répétée pour les sessions deux et trois, donnant un total de 630 tests d'accès client (70 clients * 3 accès * 3 sessions). Nous considérons deux sortes de tests d'imposture : *impos-*

tures aléatoires et impostures entraînées. Dans le cas des impostures aléatoires, les signatures CHASM des autres sujets sont utilisées pour générer les accès, donnant un total de 4830 impostures aléatoires (70 utilisateurs * 69 accès). Dans le cas des impostures entraînées, les 18 imitations disponibles pour chaque client sont utilisées, donnant un total de 1260 impostures entraînées (70 clients * 18 accès).

Avec variabilité temporelle. Le modèle du client est construit à partir des six signatures de la session un. Les tests client sont effectués sur les six signatures des sessions deux et trois, donnant un total de 840 tests d'accès clients (70 clients * 6 accès * 2 sessions). Les tests d'impostures sont effectués comme pour le protocole *sans variabilité temporelle*.

Les évaluations se font en termes de taux de Fausse Acceptation d'imposteurs FA et de taux de Faux Rejet de client FR . Ces taux varient en fonction du seuil de décision T . Les points de fonctionnement (FA, FR) sont dessinés sur un graphique (x, y) en faisant varier T . Les courbes de détection DET (Detection Error Tradeoff) utilisées dans cet article suivent ce principe en utilisant une échelle normale pour les axes x et y [MAR 97]. Les courbes DET sont généralement proches d'une ligne droite, permettant une observation plus aisée des différences de performance. Nous rapportons également nos résultats en termes de taux égaux d'erreurs (Equal Error Rate, EER) qui sont obtenus pour $FA = FR$.

4.2 Résultats

Des expériences ont été menées avec différentes tailles de modèle du monde, menant à la conclusion qu'un modèle avec 64 gaussiennes donne de bons résultats pour tous les protocoles testés. A travers les différents protocoles et en moyenne pour tous les utilisateurs, ce nombre se situe autour de 16 gaussiennes. Pour une version optimisée du système, nous pourrions envisager d'utiliser un nombre de gaussienne qui dépend de l'utilisateur, en tenant compte par exemple de la taille des signatures [LYV 03] ou en calculant une fonction de coût dépendante de la complexité du modèle et des erreurs de modélisation [RIC 03]. Pour le reste de cette section, nous rapportons des résultats pour des modèles de clients à 16 gaussiennes et le modèle du monde à 64 gaussiennes.

La figure 4 montre les courbes DET du système parole, du système signature et de la fusion des deux systèmes pour les protocoles *avec* et *sans* variabilité temporelle. De ces courbes, nous pouvons conclure que la modélisation de la parole donne de meilleures performances que la partie signature pour des tests sans variabilité temporelle. A l'inverse, lorsque les données présentent une variabilité temporelle, la partie signature donne de meilleurs résultats que la parole. La modélisation de la parole semble, pour notre système, plus sensible à la variabilité temporelle que pour la modalité signature. La fusion (avec z-norm) apporte une amélioration claire des résultats pour les deux protocoles.

Le tableau 1 résume les résultats de nos évaluations en terme d'EER pour les différents protocoles. Les conclusions suivantes peuvent être données. L'impact négatif de la variabilité temporelle est significatif dans tous les cas de figures avec une perte d'environ 15% absolus pour la modalité parole dans le cas des impostures entraînées. Une telle dégradation peut être expliquée par différents facteurs. Première-

TAB. 1 – Protocoles sans et avec variabilité temporelle, impostures aléatoires et entraînées. 16 gaussiennes sont utilisées pour les modèles de clients, 64 gaussiennes pour le modèle du monde.

variabilité temporelle	sans (%EER)		avec (%EER)	
	aléatoires	entraînées	aléatoires	entraînées
impostures				
signature	4.0	6.1	5.3	9.4
parole	2.0	3.7	14.0	19.5
fusion par somme	1.7	3.1	3.5	6.9
z-norm et fusion par som.	0.6	1.3	4.1	8.7

ment, la technique de modélisation GMM n'est sans doute pas assez robuste contre les effets de la variabilité temporelle. Une technique d'adaptation du modèle du monde pour construire le modèle du client pourrait s'avérer bénéfique à cet égard. Deuxièmement, il est également probable que le signal de parole soit naturellement plus variable que la signature. Troisièmement, la modélisation de la parole est sensible aux conditions d'acquisition (bruit ou canal), tandis que l'acquisition de la signature est plus stable par nature. L'impact négatif des impostures entraînées est également important par rapport aux impostures aléatoires. Ces résultats montrent la sensibilité de ces biométries par rapport à des attaques intentionnelles pourtant relativement simples à effectuer. La fusion apporte une amélioration claire des performances pour le protocole *avec variabilité temporelle*, ce qui tend à démontrer que l'approche CHASM est plus robuste dans des conditions plus difficiles. L'apport de la z-norm, bien que ses paramètres soient estimés a posteriori, n'est pas décisif pour la configuration actuelle de notre système.

5 Conclusions et travaux futurs

Nous avons décrit nos premiers développements vers une nouvelle méthode de vérification d'identité se basant sur une acquisition simultanée de l'écriture et de la parole. Une campagne d'acquisition de données a été menée pendant laquelle nous avons observé que tous les utilisateurs ont été capables d'enregistrer simultanément leur écriture et leur voix. Nous avons également expérimenté que les données CHASM peuvent être enregistrées suivant deux scénarii : *signatures CHASM* et *écriture CHASM*. Une enquête d'acceptabilité effectuée lors de l'enregistrement de la base de données montre que de telles acquisitions semblent acceptables pour l'utilisateur. Finalement, nous avons introduit un système de base pour la modélisation de ces données CHASM. Ce système utilise des GMMs fonctionnant sur chaque modalité séparée suivis par une simple fusion au niveau des scores. Les résultats montrent que la fusion des modalités donne globalement de meilleures performances que les modalités prises séparément. Les résultats montrent également l'impact négatif de la variabilité temporelle et des impostures entraînées. Nos futurs travaux viseront la construction de techniques de modélisation plus robustes contre la variabilité temporelle et les impostures. Dans cette optique, nous avons identifié des techniques potentielles telles que l'adaptation du modèle du monde pour créer un modèle de client, l'utilisation de HMMs ou encore la fusion au niveau des vecteurs de paramètres.

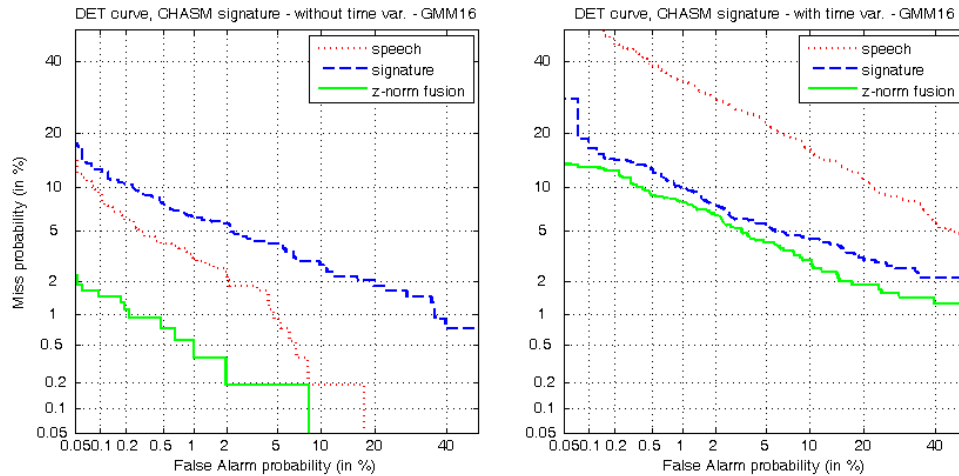


FIG. 4 – Courbes DET - fusion des systèmes GMMs parole et signature pour le protocole sans (partie gauche) et avec (partie droite) variabilité temporelle, impostures aléatoires.

Nous planifions également de consolider nos premiers résultats sur le protocole *écriture CHASM* et sur un nouveau set de données qui étendra celui utilisé dans cet article.

Remerciements : Ce travail a été principalement supporté par le Fond National Suisse pour la Recherche à travers le programme NCCR IM2 et par le projet européen BioSecure. Nous remercions vivement Asmaa El Hannani pour ses précieux conseils sur les systèmes GMM.

Références

- [AL 04] BIMBOT F. ET AL, A tutorial on text-independent speaker verification, *EURASIP Journal on Applied Signal Processing*, vol. 4, 2004, pp. 430-451.
- [AL 05] DUMAS B. ET AL, MyIdea - Multimodal Biometrics Database, Description of Acquisition Protocols, *In proc. of Third COST 275 Workshop (COST 275)*, October 27 - 28 2005, pp. 59-62, Hatfield (UK).
- [DEM 77] DEMPSTER A., LAIRD N., Maximum likelihood from incomplete data via the EM Algorithm, *Journal of Royal Statistical Society*, vol. 39, n° 1, 1977, pp. 1-38.
- [FUE 02] FUENTES M. ET AL, Identity Verification by Fusion of Biometric Data : On-Line Signature and Speech, *Proc. COST 275 Workshop on The Advent of Biometrics on the Internet*, November 2002, pp. 83-86, Rome, Italy.
- [GAR 03] GARCIA-SALICETTI S. ET AL, BIOMET : a Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities, *4th Int. Conf. AVBPA*, Springer-Verlag, 2003.
- [HEN 05] HENNEBERT J., DUMAS B., PUGIN C., EVÉQUOZ F., PETROVSKA-DELACRÉTAZ D., HUMM A., ROTZ D. V., MyIdea Multimodal Database, <http://diuf.unifr.ch/go/myidea>, 2005.
- [JAI 05] JAIN A., NANDAKUMAR K., ROSS A., Score Normalization in Multimodal Biometric Systems, *Pattern Recognition*, vol. 38, 2005, pp. 2270-2285.
- [KRA 05] KRAWCZYK S., JAIN A. K., Securing Electronic Medical Records using Biometric Authentication, *Audio- and Video-based Biometric Person Authentication (AVBPA)*, Rye Brook, NY, 2005, pp. 1110-1119.
- [LEC 94] LECLERC F., PLAMONDON R., Automatic Signature Verification : the State of the Art-1989-1993, *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 8, n° 3, 1994, pp. 643-660.
- [LEE 96] LEE L. L., BERGER T., , AVICZER E., Reliable On-Line Human Signature Verification Systems, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, n° 6, 1996, pp. 643-647.
- [LYV 03] LY-VAN B. ET AL, Signature with Text-Dependent and Text-Independent Speech for Robust Identity Verification, *Proc. Workshop MMUA*, December 2003, pp. 13-18.
- [LYV 04] LY VAN B., GARCIA-SALICETTI S., DORIZZI B., Fusion of HMM's Likelihood and Viterbi Path for On-Line Signature Verification, *Biometrics Authentication Workshop*, May 15th 2004, Prague.
- [MAR 97] MARTIN A. ET AL, The DET curve in assesment of detection task performance, *Eurospeech 1997*, Rhodes, Greece, 1997, pp. 1895-1898.
- [PLA 94] PLAMONDON R., The Design of an On-Line Signature Verification System : From Theory to Practice, *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 8, n° 3, 1994, pp. 795-811.
- [RAB 93] RABINER L., JUANG B.-H., *Fundamentals Of Speech Recognition*, Prentice Hall, 1993.
- [REY 02] REYNOLDS D., An Overview of Automatic Speaker Recognition Technology, *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 4, 2002, pp. 4072-4075.
- [RIC 03] RICHARDI J., DRYGAJLO A., Gaussian Mixture Models for On-Line Signature Verification, *Proc. 2003 ACM SIGMM workshop on Biometrics methods and applications*, 2003, pp. 115-122.
- [VIE 06] VIELHAUER C., *Biometric User Authentication for IT Security*, Springer, 2006.