



Dynamique de frappe au clavier

Projet réalisé par
Comte-Gaz Quentin,
Lillo Adrien
et Stenger Florent

Années 2012 - 2013

Sommaire

Introduction.....	3
I- La place de la dynamique de frappe (DDF) au clavier parmi les différentes biométries	4
A) Qu'est ce que la biométrie comportementale ?.....	4
B) Qu'est-ce qu'une méthode de biométrie douce ?.....	4
C) Comparaison de la DDF aux autres biométries	5
II- Types de codage de la DDF	7
A) Les méthodes statistiques	7
B) Les réseaux de neurones	8
III- Après la théorie, la pratique (expériences).....	9
A) Comparaison de la DDF de deux individus au profil similaire.....	10
B) Comparaison de la DDF entre homme et femme à âge constant	12
C) Influence de l'âge sur la DDF.....	14
D) Différenciation des enfants et des handicapés d'une main par la DDF	16
Conclusion	18

Introduction

Avec l'émergence des télécommunications et de la concurrence, la sécurité des données et l'identification des utilisateurs est devenue primordiale.

Pour ce faire, l'homme a inventé des méthodes très différentes en se basant sur la biométrie.

Nous avons centré notre étude sur l'une d'entre elles : la dynamique de frappe.

Comme nous allons le voir, cette dernière offre une identification précise tout en étant peu intrusive et performante.

Nous allons tout d'abord vous présenter la DDF et la différence qu'il peut y avoir avec les autres biométries. Nous allons ensuite vous présenter différents types de codage de la DDF et enfin vous présenter des expériences que nous avons réalisées.

La place de la Dynamique de frappe (DDF) au clavier parmi différentes biométries

La DDF au clavier est une méthode de biométrie comportementale, le plus souvent douce.

A. Qu'est-ce que la biométrie comportementale ?

La biométrie comportementale analyse chez un individu les caractéristiques de certains de leurs comportements, qui sont donc liées à des habitudes, des apprentissages et non des caractéristiques innées (par exemples : les signatures, la reconnaissance vocale).

Ce type de biométrie s'oppose aux biométries biologique et morphologique qui prennent en compte des caractéristiques innées. La première consiste en l'analyse de données telles que l'ADN ou le groupe sanguin. La deuxième utilise quant à elle des traits physiques propres aux différents individus (empreintes, rétines...).

Généralement, les techniques biologiques sont mal perçues par la population et demandent des moyens de mise en place et de fonctionnement élevés, ainsi qu'un long temps de traitement, mais présentent par contre de bonnes performances.

Les techniques morphologiques ont la plupart du temps d'aussi bonnes performances et sont mieux acceptées par le publique et plus faciles à mettre en place et à utiliser.

Finalement, concernant les techniques comportementales, elles sont réputées moins fiables, mais sont facilement acceptées dans l'opinion publique. La difficulté de leur mise en place peut varier mais est tout de même toujours bien moins compliquée que celle des techniques biologiques, et leur utilisation est extrêmement simple.

B. Qu'est-ce qu'une méthode de biométrie douce ?

Il s'agit d'une méthode biométrique qui ne demande que l'observation de l'utilisateur. C'est-à-dire que l'on va simplement analyser une caractéristique physique voyante, ou une action habituelle de la vie quotidienne de l'utilisateur pour le reconnaître.

La couleur des yeux, la forme du visage, sont donc des biométries douces, contrairement aux empreintes digitales ou de l'iris, ou encore à l'ADN, qui demandent que l'utilisateur se prête au relevé d'un échantillon.

Généralement la dynamique de frappe (DDF) est utilisée pour augmenter la sécurité lors de la saisie d'un mot de passe, ou pour vérifier au cours d'une session active que l'utilisateur est toujours le même, auquel cas il s'agit de méthodes douces.

Cependant il est aussi possible de l'utiliser pour une authentification forte, c'est-à-dire qu'au lieu d'analyser les données provenant du clavier dans son utilisation normale, il est possible

de demander à l'utilisateur de saisir une séquence clé, et de l'authentifier par l'analyse de cette séquence.

C. Comparaison de la DDF aux autres biométries

Les comparaisons précédentes étant des généralités et n'étant pas quantifiées, regardons les choses plus en détails.

Tout d'abord, le tableau suivant résume quelques résultats d'évaluations objectives (qui quantifient le système biométrique) et subjectives (qui ont trait au ressenti de l'utilisateur), pour la DDF et des méthodes biométriques usuelles ou fortement développées.

TAB. 2.1: Propriétés des modalités biométriques (le nombre d'étoiles dans la colonne performance est relié à l'efficacité de la reconnaissance)

<i>Modalité</i>	<i>Universalité</i>	<i>Unicité</i>	<i>Permanence</i>	<i>Collectabilité</i>	<i>Acceptabilité</i>	<i>Performance</i>
<i>Modalités biologiques</i>						
ADN	Oui	Oui	Oui	Faible	Faible	*****
Groupe sanguin	Oui	Non	Oui	Faible	Non	*
Signal du cerveau (EEG)	Oui	Oui	Oui	Faible	Non	****
<i>Modalités comportementales</i>						
Démarche	Oui	Non	Faible	Oui	Oui	***
Signature dynamique	Oui	Oui	Faible	Oui	Oui	****
Dynamique de frappe	Oui	Oui	Faible	Oui	Oui	***
Voix	Oui	Oui	Faible	Oui	Oui	****
<i>Modalités morphologiques</i>						
Iris	Oui	Oui	Oui	Oui	Un peu	*****
Rétine	Oui	Oui	Oui	Oui	Un peu	*****
Visage	Oui	Non	Faible	Oui	Oui	****
Géométrie de la main	Oui	Non	Oui	Oui	Oui	****
Veines de la main	Oui	Oui	Oui	Oui	Oui	*****
Oreille	Oui	Oui	Oui	Oui	Oui	*****
Empreinte digitale	Oui	Oui	Oui	Oui	Oui	****

N.B.

Universalité : tout le monde peut-il utiliser cette méthode ?

Permanence : La caractéristique évaluée reste-elle stable au cours du temps ?

Concernant la performance, il est plus aisé de se faire une idée grâce au tableau ci-dessous.

TAB. 2.2: Ordre d'idée des taux aux de performance des différentes technologies biométriques selon l'état de l'art (les valeurs peuvent changer en fonction des bases de données)

Modalité	EER
Signal du cerveau (EER)	16 - 28%
Démarche	19 - 37%
Voix	5%
Dynamique de frappe	5%
Signature dynamique	5%
Empreinte digitale	2%
Veine de la main	1.15%
Iris	≈ 0%
Visage	5 - 10%

Ces deux tableaux nous montrent que la DDF est une biométrie qu'il est tout à fait envisageable de mettre en pratique, car bien que son efficacité soit un peu inférieure à celle des systèmes morphologiques majeurs, les coûts y étant associées sont moindres. Cette technologie est d'autant plus viable que les recherches à son sujet sont encore bien peu développées par rapport à celles concernant l'iris, la voix ou les empreintes, ce qui laisse à penser que de nouveaux progrès sont à venir.

La permanence n'est pas un problème si les échantillons de références « s'ajustent » aux changements au fil du temps puisqu'il s'agit de changements très lents.

D'autre part, en plus des évaluations objectives et subjectives, il est bon de prendre en compte des critères d'évaluation sur la sécurité du système biométrique, et bien que la DDF soit possible à pirater, son utilisation est déjà en soit un gain de sécurité quand elle est ajoutée à l'identification par mot de passe.

Types de codage de la DDF

Il existe plusieurs méthodes de codages de la DDF différentes, dont font partie la méthode statistique et la méthode par réseau de neurones.

A. Les méthodes statistiques

Elles se basent sur une base de données contenant une certaine quantité d'échantillons. On peut facilement se représenter comment fonctionnent ces méthodes.

Les plus simples méthodes utilisées pour reconnaître un individu vont puiser dans un échantillon de vecteurs déjà associés à cet individu, en tirer un vecteur moyen, et y comparer le vecteur obtenu lors de la tentative d'authentification. La distance de ce vecteur au vecteur moyen va donner directement un score, qui sera utilisé pour confirmer ou non l'authentification.

Pour complexifier un peu la méthode d'authentification on peut rajouter à la précédente la prise en compte de l'écart type des vecteurs liés à un individu. Ainsi le score ne sera plus attribué linéairement à partir de la distance au vecteur moyen, mais privilégiera bien plus les vecteurs dans l'écart type par rapport à ceux en dehors. Affinée comme il faut cette variante peut améliorer le taux de discrimination et donc le FAR. On peut aussi l'adapter pour améliorer le compromis du FRR par rapport à FAR et obtenir un meilleur EER.

En plus des interrogations concernant le type de traitement, le calcul du score et les taux d'acceptation, cette méthode pose la question suivante : combien d'échantillons doit on prendre en compte ?

En effet, on parle ici de saisie au clavier, et suivant si une personne utilise son ordinateur souvent ou rarement, si elle vieillit, sa dextérité peut évoluer. Il faut donc trouver une limite acceptable au nombre d'échantillons que l'on prend en compte, et veiller à les remplacer pour ne pas en utiliser de trop anciens, sans pour autant remplacer aveuglément (il est intéressant de filtrer les vecteurs qui sont assez éloignés du vecteur moyen ; par exemple, l'utilisateur pouvant voir ses performances au clavier réduites ponctuellement, la prise en compte d'un tel échantillon pourrait diminuer l'efficacité de l'authentification à l'avenir).

B. Les réseaux de neurones

Les réseaux de neurones sont des systèmes capables d'apprendre, ils mettent en œuvre le principe d'apprentissage par l'expérience. Par confrontation avec des situations ponctuelles, ils infèrent un système de décision intégré dont le caractère générique est fonction du nombre de cas d'apprentissages rencontrés et de leur complexité par rapport à la complexité du problème à résoudre. Ainsi cette méthode ajoute une forme d'intelligence artificielle à nos algorithmes d'identification.

Les réseaux de neurones ont été utilisés dans différentes études de DDF. Ils nécessitent habituellement une quantité d'exemples d'apprentissage pour créer la référence biométrique. Le temps de calcul nécessaire pour vérifier une requête à une référence biométrique est relativement similaire pour toutes les méthodes, cependant ce n'est pas le cas pour la génération de la référence biométrique.

TAB. 3.14: Temps de calcul nécessaire pour générer les références biométriques

Nb	STAT1	STAT2	STAT3	NEURAL
5	3s	3s	3s	5m 55s
10	3s	3s	3s	30m 4s

Le tableau ci-dessus met en évidence la lenteur des réseaux de neurones relativement aux méthodes statistiques. Néanmoins, aujourd'hui on combine les deux technologies. En effet, tandis que les méthodes statistiques sont utilisées pour l'identification, les réseaux de neurones servent à la classification des utilisateurs.

Maintenant que nous avons défini les termes importants et que nous avons vu les différentes méthodes de codage de la DDF, place à la pratique.

Après la théorie, la pratique (expériences)

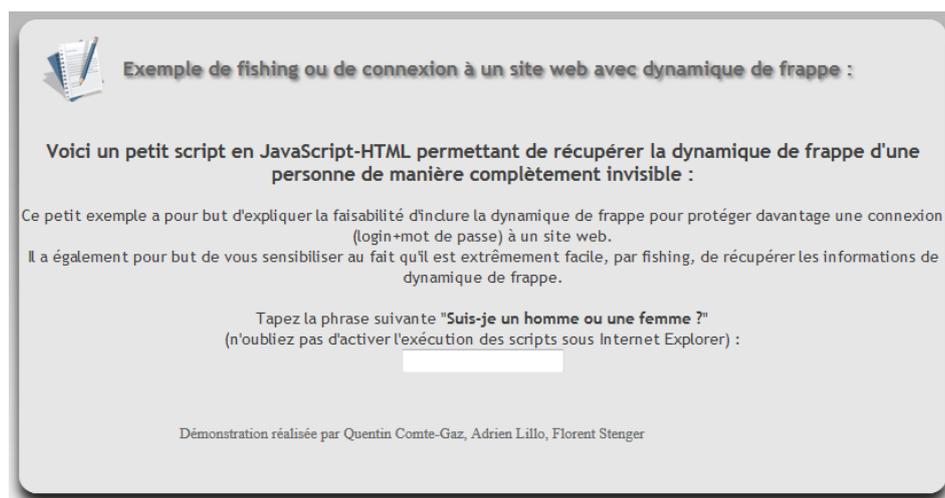
Dans cette partie, le but est de récupérer les informations de base de la DDF sur un clavier d'ordinateur (temps entre 2 touches et durée d'appui d'une touche).

Pour cela, nous avons réalisé un petit script en Javascript-HTML.

Ce script se trouve à l'adresse suivante : <http://quentincg.fr/ensea/DynamiqueDeFrappe.html>

Pour information, le Javascript est parfait pour ce mini-projet car il permet de créer directement des évènements lorsqu'une personne commence à appuyer sur une touche (onKeyDown) et à la fin de l'appui de cette même touche (onKeyUp). Nous utilisons également les fonctions getMilliseconds() et getSeconds() pour récupérer les différents temps (entre 2 touches et la durée d'appui).

Ce script réagit cependant assez mal si les temps entre 2 appuis sont très courts (<20ms), ce qui n'arrive que pour les rares personnes vraiment très rapide pour taper au clavier. En effet, la moyenne du temps entre 2 appuis est d'environ 200ms).



Impression d'écran de la page de test visualisée par l'internaute

```
function KeyDown() {  
    var date3=new Date();  
    j++;  
    debutTouche[j]=date3.getMilliseconds()+1000*date3.getSeconds();  
}  
  
function KeyUp() {  
    var date4=new Date();  
    k++;  
    finTouche[k]=date4.getMilliseconds()+1000*date4.getSeconds();  
}
```

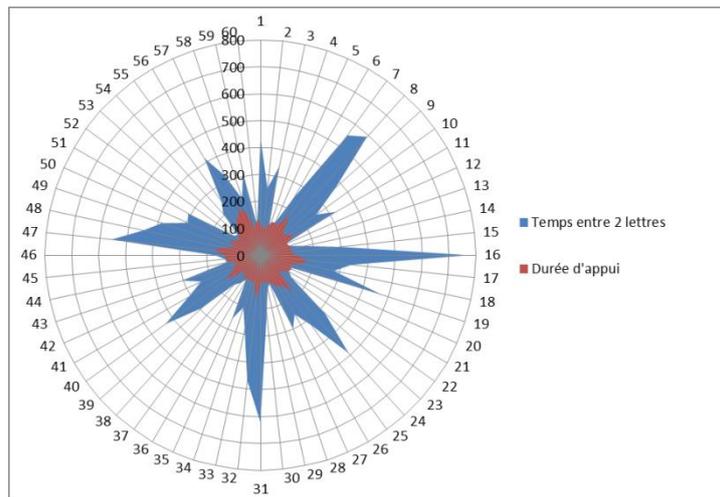
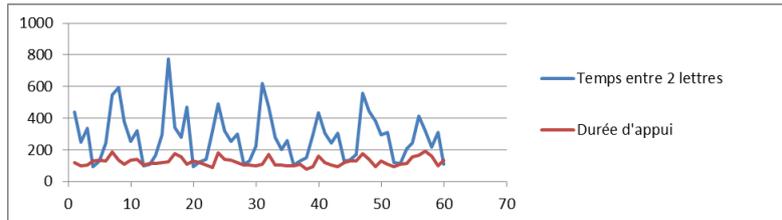
Les 2 fonctions les plus importantes du code Javascript

A. Comparaison de la DDF de deux individus au profil similaire

Le but de cette étude est de prouver que, même des personnes ayant un profil très similaire, il est possible de les distinguer grâce à la DDF.

Ainsi, pour cette étude, nous prenons deux hommes âgés de 22 ans et ayant l'habitude d'utiliser un ordinateur. Appelons les « A » et « B ».

L'utilisateur « A » a tapé plusieurs fois «bonjour,» (8 caractères), voici le résultat obtenu :



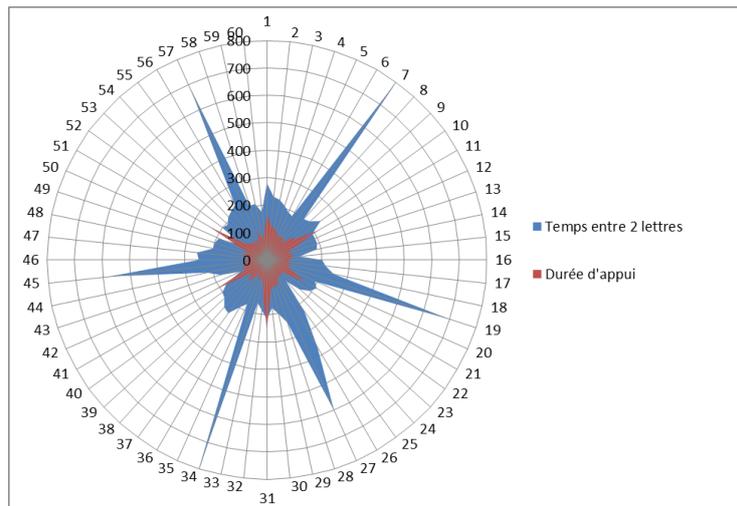
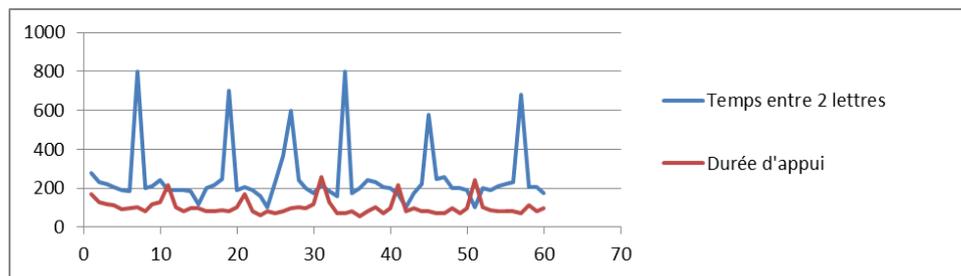
Nous pouvons noter pour cet individu que :

- Pour le temps entre 2 lettres : 2 pics et un minimum qui se repètent à chaque fois tous les 8 caractères.
- Pour la durée d'appui : C'est un peu plus confus, mais on retrouve un pic ainsi qu'un minimum pour chaque période de 8 caractères.
- Le temps moyen d'appui est de 127ms et le temps moyen entre 2 caractères est de 282ms.

Ainsi, on remarque que la durée d'appui et le temps entre 2 lettres dépendent énormément de la lettre tapée ainsi que de la lettre suivante.

On remarque également, que notre façon de taper une chaîne de caractères à deux moments différents ne varie pas énormément (on retrouve bien périodiquement la forme globale de la façon de taper).

Demandons maintenant à l'utilisateur « B » de taper les mêmes caractères, voici le resultat obtenu :



Nous remarquons ici que la personne semble (d'après le 1^{er} graphe) avoir une vitesse d'écriture assez constante sauf pour un caractère (la virgule) d'où un pic très élevé pour les 2 courbes mais un ensemble assez constant.

Nous pouvons également remarquer un petit soucis : la personne qui a tapée les caractères a, de temps en temps, ajouté un espace (pour une raison obscure). Cela a eu pour conséquence d'avoir, de temps en temps, des périodes de 9 caractères au lieu de 8. Les 4 espaces indésirables se remarquent très nettement : Nous observons 4 minimums pour la courbe de temps entre 2 caractères.

En négligeant les espaces, nous obtenons un temps moyen d'appui de 92ms et un temps moyen entre 2 caractères de 220ms.

Les temps moyens sont assez similaires entre les 2 individus assez similaires alors que la manière d'écrire est totalement différente.

La manière de taper des caractères au clavier dépend des individus. Ainsi, pour identifier un individu parmi un grand nombre de personnes, il est préférable d'étudier les caractéristiques propres à l'individu telles que les maximums et minimums locaux, la constance dans l'écriture...

B. Comparaison de la DDF entre homme et femme à âge constant

D'après trois chercheurs (Romain Giot, Bernadette Dorizzi et Christophe Rosenberger), il serait possible, de détecter le sexe d'un individu à partir de sa dynamique de frappe.

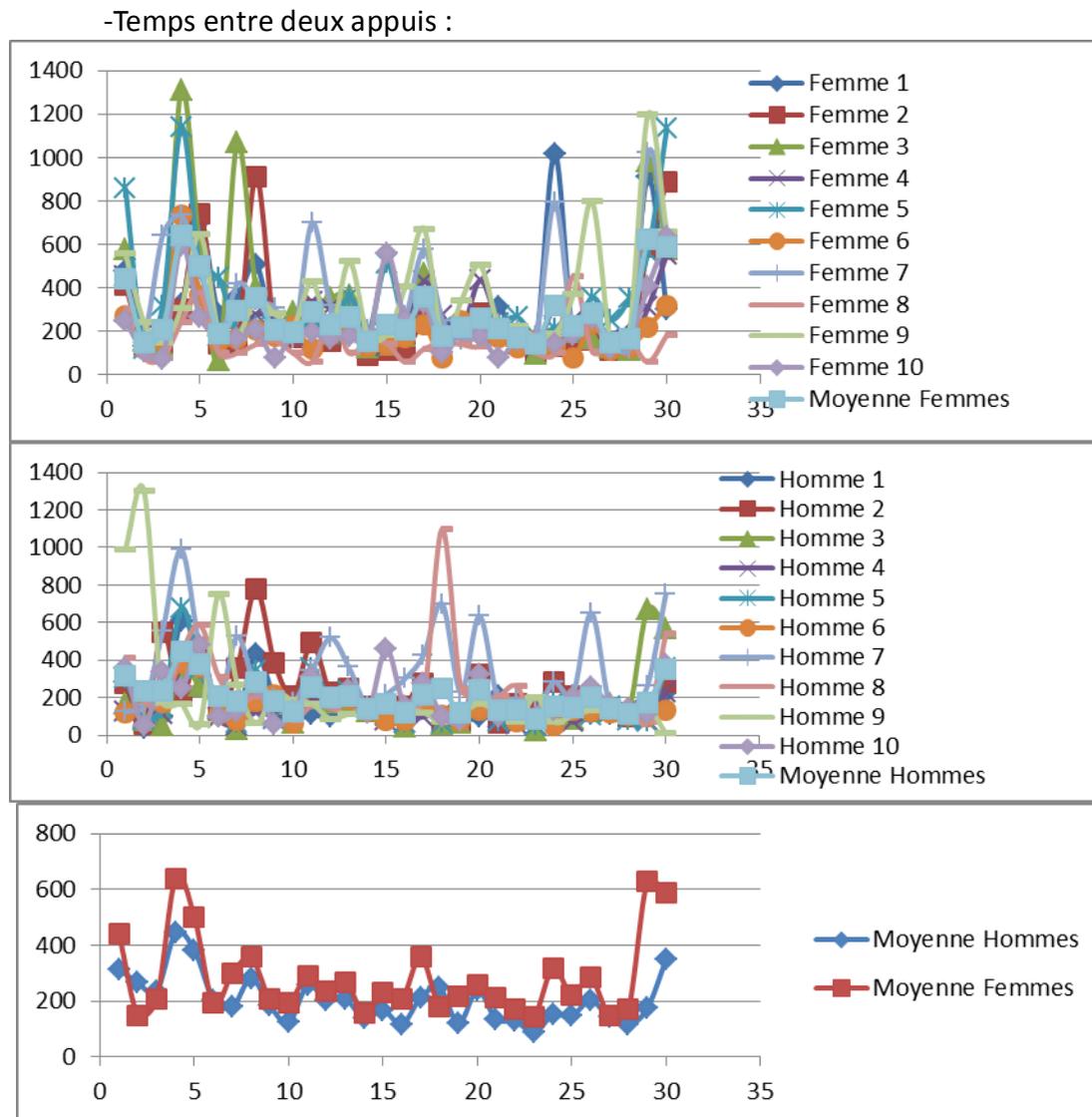
Nous allons donc tenter l'expérience pour nous en assurer :

Pour normaliser l'expérience, nous allons nous limiter à des personnes ayant entre 20 ans et 25 ans et leur faire taper une phrase type (« Suis-je un homme ou une femme ? »).

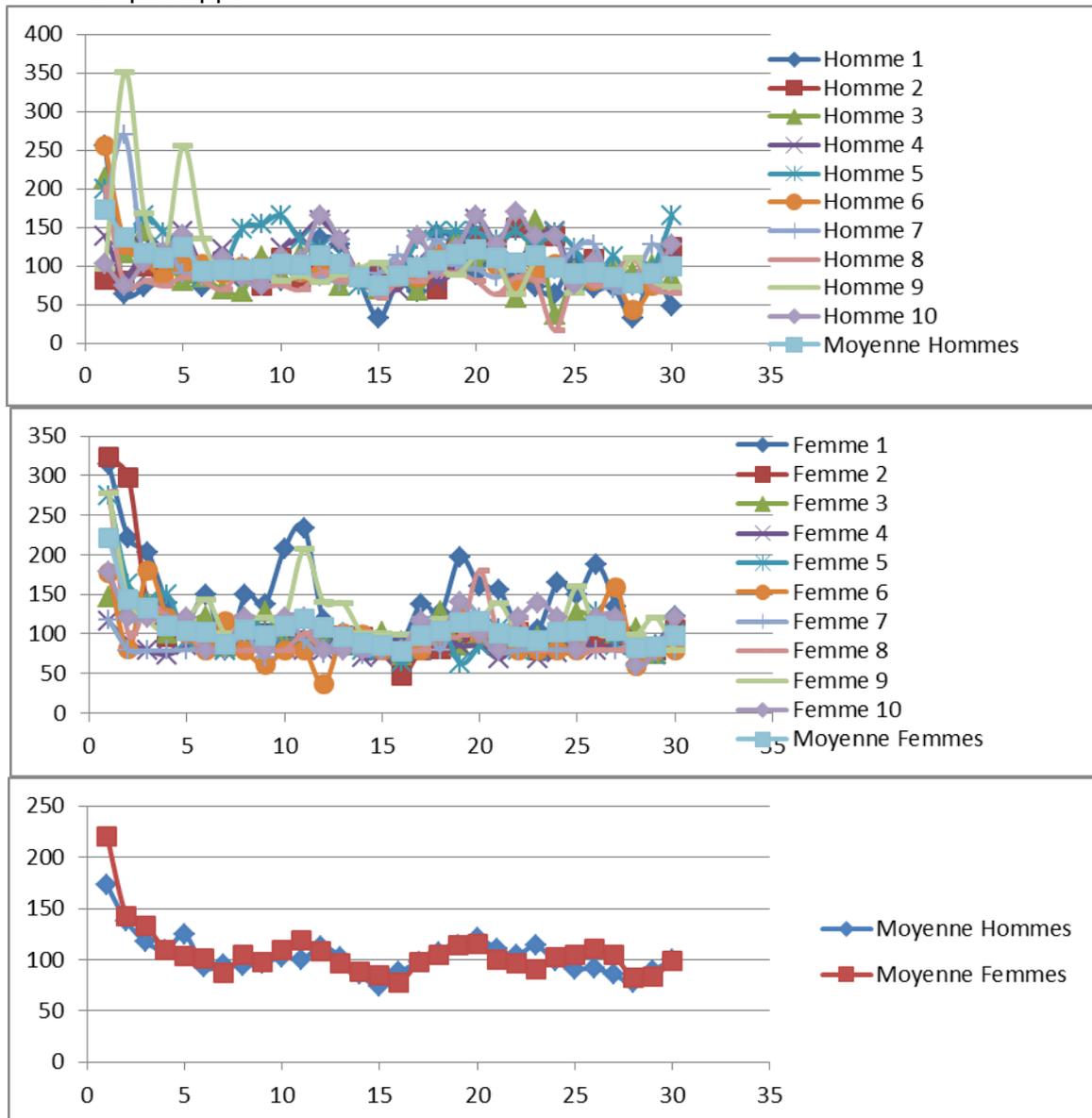
Nous allons donc prendre un échantillon de 10 femmes et 10 hommes âgés d'environ 22 ans (ne pouvant pas avoir un échantillon plus grand, nous avons préféré nous focaliser sur une catégorie de population bien particulière).

Le but étant de déterminer s'il y a vraiment une différence dans la manière d'écrire entre les hommes et les femmes.

Voici les résultats obtenus :



-Temps d'appui :



Cela semble assez peu prometteur. En effet, les hommes et les femmes semblent avoir un comportement assez similaire.

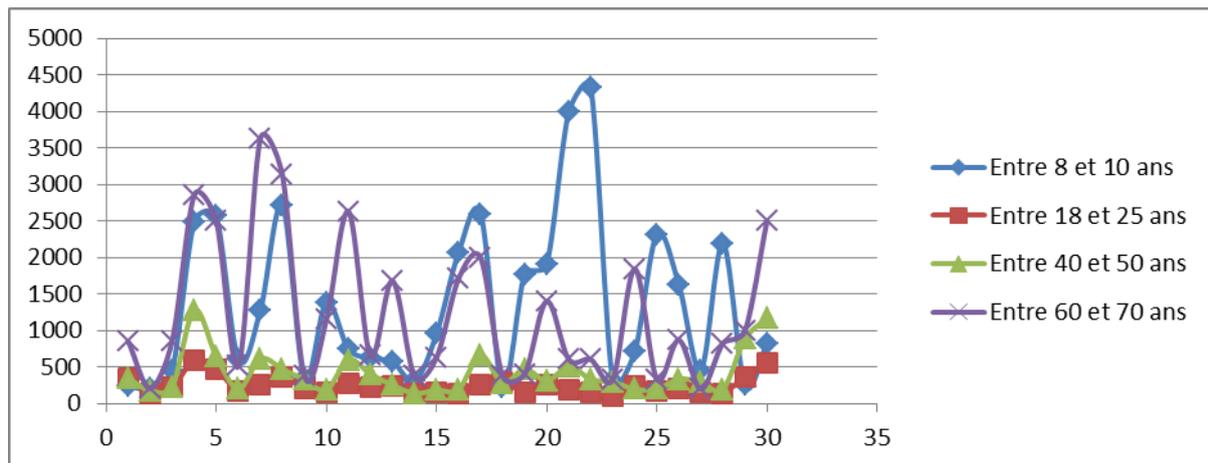
Nous avons donc essayé, au lieu de faire une simple moyenne, de passer à la variance. La variance nous fait remarquer que les hommes semblent avoir une constance dans la vitesse de frappe et dans le temps d'appui. Les femmes semblent un peu moins constantes. Cela n'est pas flagrant mais, sur les 10 plus constants dans la façon de taper au clavier, 8 sont des hommes. Ceci est bien entendu à vérifier sur un groupe de personnes beaucoup plus large.

La variance pour les femmes est environ 1.2 fois plus importante que chez les hommes. Pour être encore plus sûr que la personne est vraiment un homme ou une femme, il faudrait analyser la variance sur un texte plus long. En effet, ici le texte est très court et ne nous permet pas de faire des analyses très poussées des résultats obtenus. Nous pourrions également étudier les variations de manière de taper une même lettre. Beaucoup d'analyses sont possibles, nous n'avons malheureusement pas le temps de pousser nos recherches.

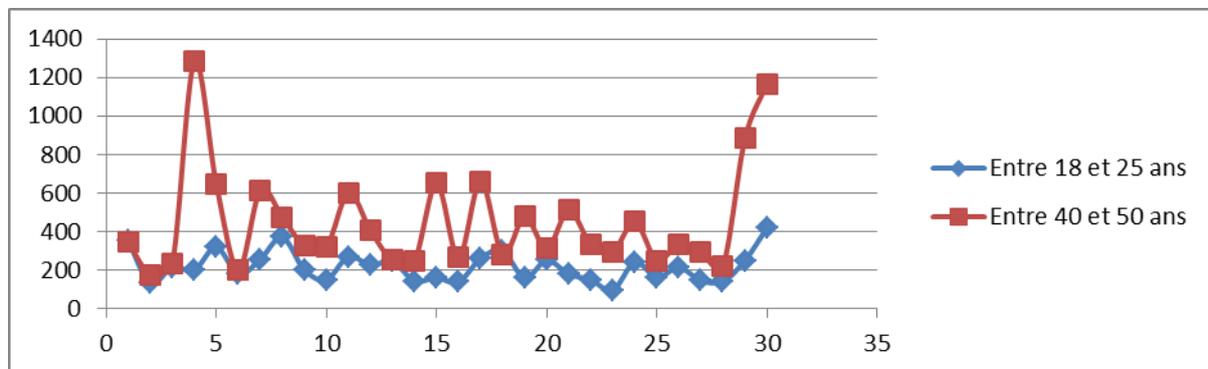
C. Influence de l'âge sur la DDF

Pour cette expérience, nous avons fait passer le test précédent (faire taper la phrase type « Suis-je un homme ou une femme ? ») à des personnes plus âgées et plus jeunes. Pour simplifier la visualisation des courbes, nous avons fait la moyenne par groupe d'âge : entre 8 et 10 ans (3 individus), entre 18 et 25 ans (15 individus), entre 40 et 50 ans (6 individus) et enfin entre 60 et 70 ans (1 individu).

Voici les différents graphes obtenus :



Temps entre 2 touches (1)

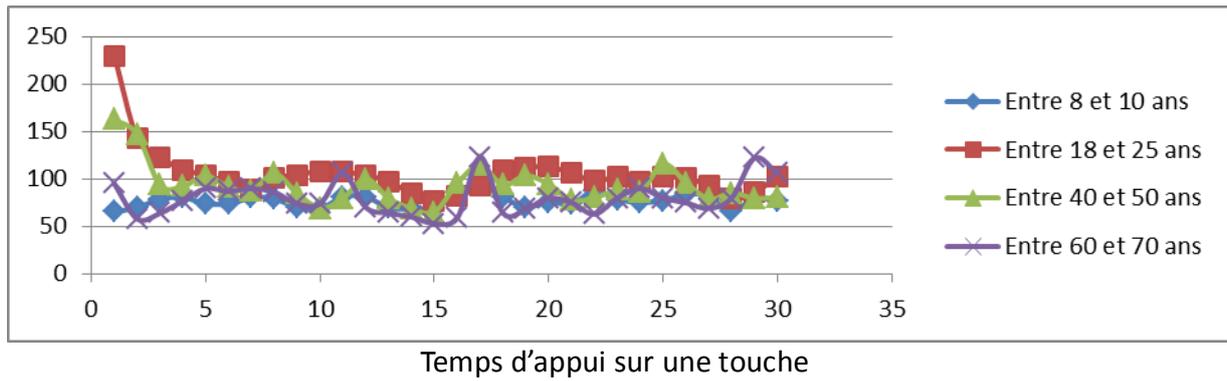


Temps entre 2 touches (2)

On remarque que les enfants et les personnes âgées sont très lents pour trouver les différentes touches. Cela vient sans doute du fait qu'ils ne sont pas encore habitués à taper au clavier.

On remarque ensuite que les personnes ayant entre 40 et 50 ans sont assez rapides pour trouver les différentes lettres. Mais les plus rapides restent bien ceux qui ont nagé dans l'informatique depuis la naissance (les 18-25 ans).

Le temps d'appui sur une touche ne semble par contre ne pas être un critère prépondérant pour déterminer l'âge d'une personne :



En conclusion, avec une simple moyenne du temps entre touches, on peut connaître approximativement l'âge d'une personne avec taux d'erreur « assez faible ». Il est cependant difficile avec une simple moyenne de distinguer les enfants des personnes ayant plus de 60 ans. Pour trouver des différences entre ces 2 catégories, il va falloir analyser la des caractéristiques particulières (des minimums et des maximums sur certaines touches).

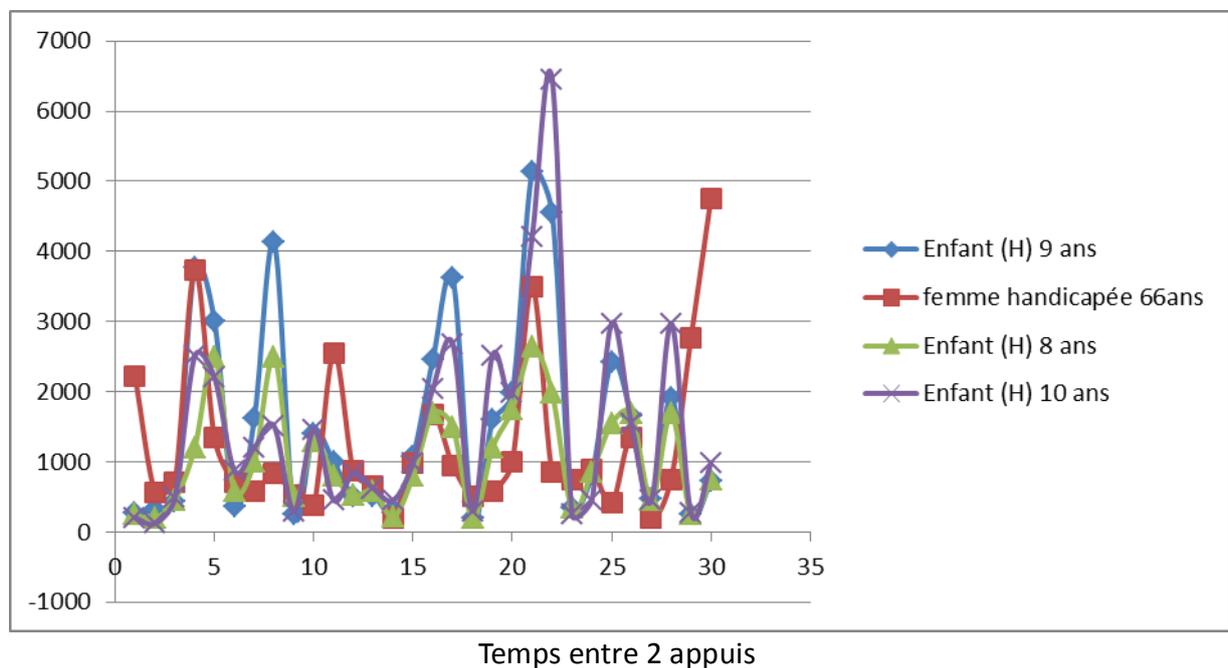
D. Différentiation des enfants et des handicapés d'une main par la DDF

Reconnaitre un enfant ou une personne handicapée parmi une foule de personnes d'âge moyen est assez facile (le temps entre deux touches est bien plus faible chez les personnes d'âge moyen).

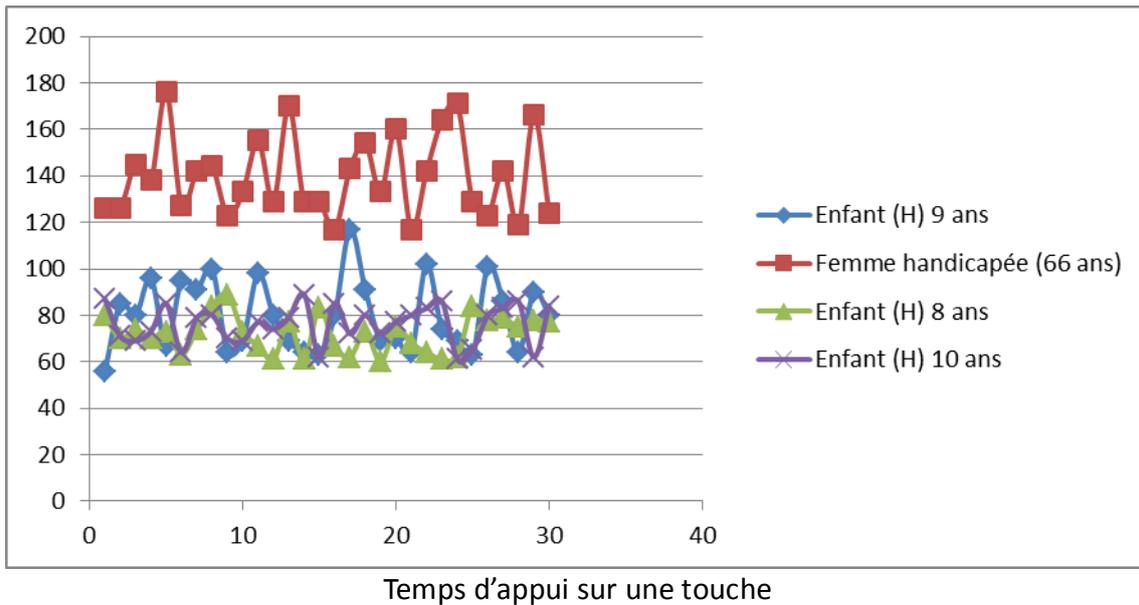
Dans cette section, nous allons essayer de faire quelque chose de plus délicat : distinguer la DDF d'une personne handicapée, d'au moins une main, d'un enfant.

Au premier abord, cela semble compliqué car ces deux catégories sont généralement lentes à taper au clavier.

Comparons le temps entre touche et le temps d'appui d'une gauchère ayant des problèmes avec sa main gauche à ceux de trois enfants :



Il est ici assez difficile de différencier les enfants des personnes âgées avec le temps entre 2 appuis ... comment faire alors ? Le temps d'appui ne nous a pas été utile jusque là, est-ce que cela va changer ?



On remarque, avec toutes les expériences précédentes et avec le graphe ci-dessus, que la moyenne du temps d'appui sur une touche est propre à la personne. De plus, cette variable est assez constante pour un individu et varie entre les individus. D'après le graphe précédent, on remarque qu'une personne tapant d'une main va faire des appuis longs et avoir un temps entre touches significatif alors qu'une personne tapant à deux mains va faire des appuis courts et avoir un temps entre touches plus ou moins variable.

Un enfant va pour sa part avoir tendance à chercher longement les lettres mais à faire des appuis courts, au contraire des personnes tapant à une main qui vont garder les touches enfoncées plus longtemps.

Ainsi, un des paramètres (temps d'appui) qui nous paraissait assez inutile jusque là a bien son importance : Ce paramètre est propre à la personne (ce qui permet, dans une certaine mesure, de l'identifier parmi un groupe d'utilisateur) et permet surtout de remarquer des minorités de personnes (handicapés, personnes utilisant une seule main ...).

Conclusion

Après cette étude sur la DDF, nous pouvons en conclure que cette biométrie douce n'est pas parfaite. En effet, l'EER de cette méthode est non négligeable (mais acceptable) ; De plus, cette méthode ne fonctionne pas si l'on a une main de prise en même temps que l'on tape son mot de passe ou si l'on a une main blessée ...

Cependant, cette méthode est prometteuse car elle peut être installée sur un ordinateur sans aucun ajout matériel (et donc presque gratuitement).

Ainsi, une méthode d'identification sur ordinateur combinant la DDF et un mot de passe serait un pas en avant pour la sécurité des données personnelles et permettrait de lutter contre une recherche de mot de passe par dictionnaire ou/et par force brute.

Remarquons tout de même que cette méthode n'est pas destinée uniquement aux ordinateurs ; En effet, des études sur la DDF pour tablettes tactiles sont actuellement en cours.