

Identification des segments EMG par Data Mining et Statistiques

M. Atieh^{*}, R. Younès^{**}, M. Khalil^{**}, N. Matta^{*} et H. Akdag^{***}

^{*} *Université de Troyes, Laboratoire Tech-CICO, Troyes, France*

younesmi@utt.fr, nada.matta@utt.fr

^{**} *Faculté de Génie – Université Libanaise, Equipe 3M, Beyrouth, Liban*

ryounes@ul.edu.lb, mkhalil@ieee.org

^{***} *Université de Reims, Laboratoire L.E.R.I, Reims, France*

Herman.Akdad@lips.fr

Abstract: The objective of this paper is to evaluate and detect muscular tiredness in natural activities, in particular, to select the most comfortable car seat. This work consists to identifying and to classify the EMG signal of the techniques from data mining and from the statistical techniques. We thus tried hybridization between some to lead to a better separation between the classes. The methods of clustering will be applied to some signals resulting from the experiments of discomfort of long duration on seats of different vehicles. These methods consist in separating segments EMG in two classes corresponding to the frequential variation from EMG signal.

Keywords: EMG Signal, Data Mining, Clustering, Muscular tiredness.

1 Introduction

Les signaux myo-électriques, appelés aussi électromyogrammes (EMG), sont des signaux électriques pouvant être enregistrés au niveau des muscles. Ces signaux biologiques sont étudiés depuis plusieurs décennies. Plus faciles à mesurer que d'autres signaux émis par les nerfs ou bien encore au niveau du cerveau, ils peuvent fournir des informations diverses sur un individu.

L'objectif de notre étude est de détecter la fatigue musculaire à partir de l'analyse du signal électromyogramme EMG. Les applications issues de ce type d'étude sont diverses. Parmi celles-là nous pouvons citer : l'étude de la fonctionnalité des muscles et des jonctions neuromusculaires, la médecine de réadaptation, l'évaluation des moments et des forces spinaux, l'étude de l'activité musculaire en charge de levage, l'analyse de l'effet physiologique du travail... Nous nous intéressons plus à l'évaluation de la fatigue musculaire dans des applications d'ergonomie. Des essais d'acquisition des EMG ont été menés sur une personne, devant conduire pendant 2h30mn, et installées sur deux sièges d'automobiles différents.

Des études sont faites pour interpréter les signaux EMG en tant que commandes de contrôle. Une

expérience concluante [Jorgensen & al, 2000] a ainsi été menée pour faire atterrir en simulation un 757 sur l'aéroport international de San Francisco. Le pilote, au lieu d'utiliser un manche de contrôle, fermait sa main devant lui et effectuait des mouvements, capturés à l'aide d'électrodes, puis analysés. Toujours dans le domaine, Sony étudie la possibilité d'employer les EMG pour le contrôle d'un lecteur de musique à l'aide de gestes prédéfinis de la main [Dubost & al, 2002]. Un domaine d'application particulièrement important de l'utilisation des EMG est la réhabilitation des personnes handicapées. Les EMG peuvent ainsi être employés pour commander des fauteuils ou bien des prothèses de main ou de bras, appelées prothèses myo-électriques [Huet 2004, Younès & al 2005].

Après une première présentation de l'article, une deuxième partie décrit le signal EMG et ses divers paramètres. Le troisième paragraphe représente une description des différentes méthodes utilisées pour la classification des signaux suivant des techniques de Data mining et des techniques statistiques. Cette partie est suivie par la présentation des applications faites à ce stade et les résultats obtenus. Avant de terminer par la conclusion et les perspectives de ce domaine d'application, nous avons effectué une comparaison entre les différentes méthodes suivant des critères d'évaluation présentés dans le paragraphe.

2 Méthodologie

2.1 Définition du signal EMG

Le signal électromyogramme (EMG) est une addition des trains de potentiel d'action des unités motrices qui sont détectés par un système d'électrodes à proximité des fibres. Quand les électrodes sont placées sur la surface de la peau, le signal détecté est désigné sous le nom d'électromyogramme de surface (SEMG). L'amplitude du signal EMG est stochastique (aléatoire) avec une distribution gaussienne qui s'étend de 0 à 6 mV (crête à la crête). L'énergie utilisable du signal est limitée (de 0 à 500 hertz de gamme de fréquence, avec l'énergie dominante étant dans la gamme de 50-150 hertz) [El Falou & al 2002].

Chaque signal acquis est composé de 70.000 points. Il sera divisé en segments dont chacun est considéré en état statique. Nous devons calculer, pour chaque segment, l'ensemble des paramètres cités ci-dessous. Les méthodes de classification serviront à trouver, dans chaque matrice de paramètres, deux classes distinctes: muscle fatigué ou non fatigué.

2.2 Paramètres du signal

Nous avons étudié les deux types classiques de paramètres statistiques et spectraux [khalil 2002].

2.2.2 Paramètres spectraux:

<p>Les fractiles f_k :</p> $\int_{p-1}^{f_k} S_x(f)df = k \int_0^{f_{\max}} S_x(f)df$ <p>$0 < k \leq 1$</p>	<p>Moments spectraux :</p> $M_r = 2 \int_0^{+\infty} f^r S_x(f)df$
<p>L'énergie par bande de fréquence :</p> $W_n = \frac{\int_{f_{n-1}}^{f_n} S_x(f)df}{M_0}$ <p>avec $f_n = \frac{n}{N} f_{\max}$ et $1 < n < N$</p>	<p>La fréquence Médiane :</p> $\int_0^{f_{med}} S_x(f)df = \int_{f_{med}}^{f_{\max}} S_x(f)df$
<p>Le rapport H/L (High/Low) :</p> $\frac{H}{L} = \frac{\int_{f_{H2}}^{f_{H1}} S_x(f)df}{\int_{f_{L1}}^{f_{L2}} S_x(f)df}$ <p>$H = [H_1, H_2]$ et $L = [L_1, L_2]$</p>	<p>La fréquence moyenne:</p> $MPF = M_1 / M_0$

Table 1 : Paramètres spectraux étudiés

2.2.1 Paramètres statistiques:

<p>La valeur moyenne :</p> $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	<p>l'écart type :</p> $\sigma = \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}$
<p>La dissymétrie :</p> $y = \frac{E(x - \bar{x})^3}{\sigma^3}$	<p>L'aplatissement :</p> $k = \frac{E(x - \bar{x})^4}{\sigma^4}$

Table 2 : Paramètres statistiques étudiés

3. Description des méthodes de classification

Nous avons procédé de différentes façons. Nous avons essayé certaines techniques statistiques (k-moyenne, LBG, ACP) ainsi que d'autres de Data mining (la logique floue, les réseaux compétitifs, algorithme des fourmis), pour résoudre le problème de classification des segments de chacun des signaux, sujets de notre étude.

3.1 Méthodes statistiques :

3.1.1 Algorithme des K-Moyennes :

Son principe est le suivant: « on dispose de points de l'espace des observations que l'on souhaite rassembler en classes, sans que l'on dispose de connaissance a priori de propriété(s) particulière(s) sur ces classes ; seul leur nombre p est fixé a priori ». Son algorithme est [Dreyfus, 2002]:

- 1) définir K **centroïdes** initialisés au hasard.
- 2) « attribuer » chaque **point** au centroïde le plus proche.
- 3) Calculer les coordonnées de chaque centroïde comme la moyenne des points qui lui sont attribués.
- 4) Tant que les centroïdes évoluent, aller en 2.

3.1.2 Algorithme LBG:

L'algorithme des K-moyennes n'étant que localement optimal, le choix des noyaux initiaux est important. Une variante proposée par Linde, Buzo Gray [Linde & al 1980], consiste à procéder hiérarchiquement et réalise une sorte d'initialisation itérative : il s'agit de choisir un seul centroïde (centre de gravité des éléments), l'éclater en deux puis chacun converge vers un ensemble de points.

3.1.3 Analyse en composantes principales:

L'ACP est une méthode classique utilisée pour la réduction de dimension : elle permet d'extraire, des données d'un nuage de points multidimensionnel, les directions importantes. Soient S1, S2 ...Sn les paramètres des signaux à classifier; chaque signal est constitué de Np points. Pour l'analyse en composantes principales, on fait l'approximation que ces Np points forment une ellipsoïde à n dimensions ce qui permet de calculer: (a) les axes de cet ellipsoïde qui forment la base ACP ; (b) la longueur de chacun d'eux.

Ces deux valeurs sont particulièrement intéressantes car elles permettent, d'une part, d'identifier la direction principale de l'ellipsoïde, et de quantifier sa taille et son volume d'autre part. On cherche donc ici les coordonnées de ces Np points dans la base ACP (P1, P2, ..., Pn); ce qui nécessite le calcul de la matrice de passage de la base (S1, S2, ..., Sn) à la base (P1, P2, ..., Pn). La matrice de passage s'obtient par diagonalisation de la matrice de covariance C associée aux Np points.

On calcule maintenant D la matrice diagonale des valeurs propres et λ et P la matrice de passage de la base (P1, P2, ..., Pn) à la base (S1, S2, ..., Sn) constituée des vecteurs propres de C; Le signal (S1, S2, ..., Sn)

dans cette nouvelle base (P_1, P_2, \dots, P_n) s'écrit donc (ayant pris soin de classer les valeurs propres par ordre croissant) : $[S_1^{ACP}, S_2^{ACP}, \dots, S_n^{ACP}] = P^{-1}[S_1 \ S_2 \ \dots \ S_n]^T$

S_1^{ACP} est la voie qui possède la plus grande dynamique [Saporta 1999].

3.2- Méthodes de Data mining:

3.2.1 Algorithme de K-moyenne flou :

L'algorithme des K-moyennes floues (fuzzy c-means) est un algorithme de classification floue fondé sur l'optimisation d'un critère quadratique de classification où chaque classe est représentée par son centre de gravité. Cette méthode nécessite de connaître le nombre de classes au préalable et génère les classes par un processus itératif en minimisant une fonction objective.

L'algorithme utilise l'ensemble des paramètres $A = \{x_1, x_2, \dots, x_n\}$ où x_i est un vecteur de plusieurs composantes, et le nombre de région c . Les valeurs des degrés d'appartenance sont regroupées dans une matrice $U = [u_{ik}]$ pour : $1 \leq i \leq n, 1 \leq k \leq c$, où u_{ik} désigne le degré d'appartenance du segment i à la classe k . Pour avoir une bonne partition, on impose aux éléments de U les contraintes suivantes :

$$u_{ik} \in [0,1]; \sum_k u_{ik} = 1; \forall i$$

L'algorithme du FCM fait évoluer la partition (Matrice U) en minimisant la fonction objective suivante :

$$J_m(U, c) = \sum_i \sum_k (u_{ik})^m \|x_i - c_k\|^2$$

où : $m > 1$ est un paramètre contrôlant le degré de flou (généralement $m = 2$) ; c_k : le centre de la classe k .

3.2.2 Réseaux de neurones compétitifs:

Le réseau compétitif permet aux neurones de sortie de se concurrencer afin qu'un seul neurone soit activé. Pour le neurone gagnant, le niveau de son activité interne, pour un signal d'entrée, doit être le plus grand par rapport aux autres neurones du réseau. Le signal de sortie du neurone gagnant sera égal à 1, alors que le signal de sortie du neurone perdant sera égal à 0. Par conséquent, chaque neurone d'un réseau apprend à se spécialiser sur un ensemble de signaux d'entrée et devient un détecteur spécialiste.

Dans sa forme la plus simple, le réseau compétitif possède un réseau de neurones à une seule couche de neurones de sortie, chacun étant complètement connecté aux signaux d'entrée [Dreyfus, 2002].

Le réseau peut inclure des connexions latérales de poids fixes entre les neurones permettant d'effectuer des inhibitions latérales tel que chaque neurone tend à inhiber le neurone auquel il est relié. Si un neurone particulier gagne la compétition, alors chaque entrée de ce neurone gagnant change son poids tout en gardant une somme unité. L'ajustement appliqué au poids synaptique sera:

$$\Delta_{w_{kj}} = \begin{cases} \frac{1}{0}^{(x_j - w_{kj})} & \text{si le neurone } k \text{ est gagnant} \\ 0 & \text{si le neurone } k \text{ est perdant} \end{cases}$$

Et alors, chaque neurone sera spécialiste d'un groupe d'échantillons se présentant à l'entrée du réseau. A la présentation d'un vecteur d'entrée, un neurone de sortie est sélectionné. Il correspond le plus possible à cette entrée (minimisation d'une distance). La loi de modification des poids des connexions est dérivée de celle de Hebb.

3.2.3 Algorithme des fourmis:

Pour rassembler en tas un ensemble d'objets de même type par des fourmis réelles, les probabilités de ramasser un objet (P_p) et de le déposer (P_d) ont été explicitées:

$$P_p = \left(\frac{k_1}{k_1 + f} \right)^2 \quad P_d = \left(\frac{f}{k_2 + f} \right)^2$$

où k_1 et k_2 sont des constantes positives et f correspond au nombre d'objets rencontrés durant les T derniers déplacements divisé par le nombre maximum d'objets qui auraient pu être rencontrés. Quand il y a peu d'objets dans le voisinage de l'objet convoité par la fourmi, $f \ll k_1$ ce qui signifie que P_p est proche de 1 et l'objet a beaucoup de chance d'être ramassé. Inversement, quand le voisinage est dense en éléments, $f \gg k_1$ et alors P_p est proche de 0.

Les résultats obtenus par simulation montrent l'apparition de groupes d'objets. Les agents ainsi définis permettent donc de ranger une surface sur laquelle des objets ont été éparpillés. Le comportement de rassemblement se transforme alors en tri, ce qui se rapproche de notre problème de partitionnement [Monmarché 2000].

Lumer et Faieta ont proposé un algorithme utilisant une mesure de dissimilarité entre les objets [Lumer & al, 1994]. Les objets qui, rappelons-le, correspondent à des points d'un espace numérique à N dimensions sont plongés dans un espace discret de dimension moindre (typiquement de dimension 2). Cet espace discret s'apparente alors à une grille G dont chaque case peut contenir un objet. Les agents se déplacent sur G et perçoivent une région R_s de $s \times s$ cases dans leur voisinage. La fonction de densité locale est calculée de la manière suivante :

$$f(o_i) = \begin{cases} \frac{1}{a^2} \sum_{o_j \in R_s(o_i)} \frac{d(o_i - o_j)}{a} & \text{si } f \neq 0 \\ 0 & \text{sin on} \end{cases}$$

Comme on peut le remarquer, $f(o_i)$ dépend de l'objet considéré o_i et de sa position sur la grille $r(o_i)$. $f(o_i)$ est alors une mesure de la similarité moyenne de l'objet o_i avec les objets o_j présents dans son voisinage. a est un facteur d'échelle déterminant dans quelle mesure la dissimilarité entre deux objets est prise en compte.

4. Application des méthodes de classification

Ces différentes méthodes ont été appliquées sur deux signaux réels “EMGX” (figure 1.) et “EMGY” (figure 2.) issus d’une même personne testée sur deux types de sièges « X » et « Y » à confort différent. L’analyse de ces signaux doit nous permettre de choisir le siège le plus confortable.

4.1 Choix des données pertinents

Le signal, tel qu’il a été acquis, ne peut être lu qu’à partir ses paramètres. D’où la nécessité de calculer les différents paramètres du signal. Toutefois, le choix de ces paramètres doit prendre en compte les critères suivants [Huet 2004] :

Une séparation maximale des classes : un espace de caractéristiques est de bonne qualité si les résultats y sont classés dans des ensembles qui ont une séparation inter-classe maximale et un chevauchement minimum. Ce critère permet de s’assurer que le taux d’erreur de classification résultant soit le plus faible possible.

La robustesse : le choix de caractéristiques sélectionnées doit être valide même en changeant le signal.

La complexité : la complexité de calcul des paramètres doit être restreinte de manière à ce que la procédure de l’identification des EMG puisse être implémentée avec un dispositif matériel raisonnable.

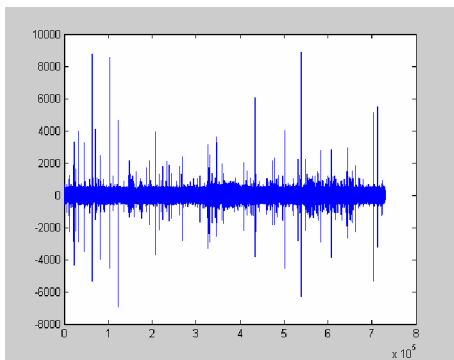


Figure 1 : Signal réel « EMGX »

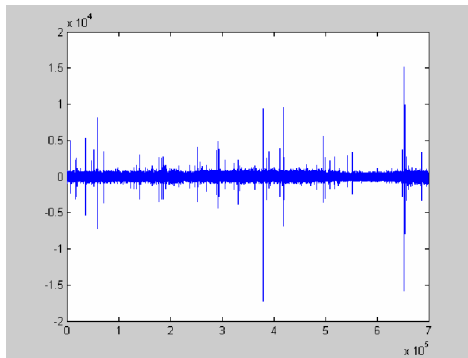


Figure 2 : Signal réel « EMGY »

Le tableau suivant représente, en détail, les taux d’erreur donnés par chaque type de paramètres, testés sur le signal « EMGY »:

Classe C1	Classe C2	Taux d’erreur
Paramètres statistiques	Pas de résultats	
M_1 et M_2	Pas de résultats	
W_n	16 %	8.3 %
f_k	20 %	28 %
F_{med} , MPF , H/L	18 %	7 %
Paramètres spectraux	Pas de résultats	
Tout sauf M_1 et M_2	12 %	27 %

Table 3 : taux d’erreur par type de paramètres

Les paramètres statistiques n’ont aucune influence sur la détection de la fatigue; Parmi les paramètres spectraux, on a remarqué que les énergies de fréquence ont donné les meilleurs résultats de classification répondant aux critères cités ci-dessus.

4.2 Résultats des K-MOY, LBG, FCM et RNC :

D’après les diverses méthodes appliquées, le premier signal « EMGX » nous a montré une faible variation du contenu fréquentiel, à partir du 81^e segment, indiquant le confort du siège « X ». Le deuxième signal indique une fatigue qui se traduit par un changement fréquentiel bien claire qui apparaît à partir du 51^e segment.

		Classe C1	Classe C2	Taux d’erreur
EMGX	K-MOY	Segments 1- 81 $\Rightarrow \geq 55\%$	Segments 82- 146 $\Rightarrow \geq 45\%$	37.7 %
	LBG			36.3 %
	FCM			35 %
	RNC			34.9 %
EMGY	K-MOY	Segments 1-50, 112-139 $\Rightarrow \geq 56\%$	Segments 51- 111 $\Rightarrow \geq 44\%$	12.3 %
	LBG			12.3 %
	FCM			12.3 %
	RNC			12.3 %

Table 4 : Résultats de classification obtenus

4.2 - Résultats de ACP :

Le nombre d’axes à retenir est défini selon le critère de Kaiser [Saporta 1999]. On préconise de détecter sur le diagramme des valeurs propres l’existence d’un coude. Le « scree-test » de Cattell [Saporta 1999] en est la version analytique. Il consiste à :

(1) calculer les différences premières des valeurs propres $\lambda_1 - \lambda_2 = \varepsilon_1$ et $\lambda_2 - \lambda_3 = \varepsilon_2 \dots$;

(2) puis calculer les différences secondes $\varepsilon_1 - \varepsilon_2 = \delta_1$ et $\varepsilon_2 - \varepsilon_3 = \delta_2 \dots$;

(3) on retient alors les valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_k, \lambda_{k+1}$, telles que $\delta_1, \delta_2, \delta_k$ soient tous positifs.

L’application de ce critère sur les signaux

« EMGY » et « EMGX » donne les résultats suivants :

	Signal « EMGX »	Signal « EMGY »
Les différences secondes	0.0056	0.0061
	0.0024	0.0003
	0.0014	-0.0001
	0.0002	0.0001
	0.0000	0.0004
	0.0001	0.0000
	0.0000	0.0000
	0.0000	0.0000

Table 5 : Application de Critère de Catell

On remarque bien que deux voies sont suffisantes pour «EMGY» et quatre pour «EMGX». Cependant, ce critère n'est pas absolu; il donne des résultats suffisants mais pas obligatoirement nécessaires. Il se peut que le nombre de voies utile pour la séparation en deux classes soit inférieur au nombre donné par le critère de Cattell, comme dans le cas de «EMGY» où, d'après les figure 3 et 4, une seule voie est suffisante et donc la classification sera détectée visuellement, ce qui n'est pas le cas de «EMGX».

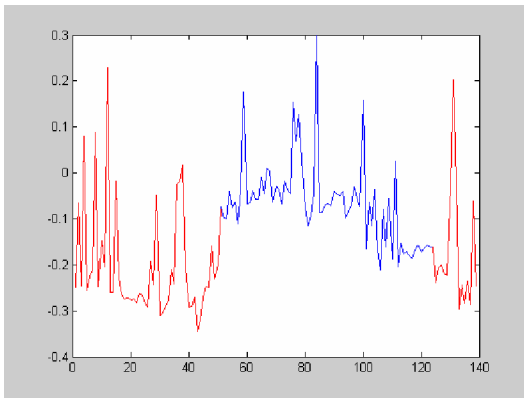


Figure 3 : Voie ACP du signal réel «EMGY»

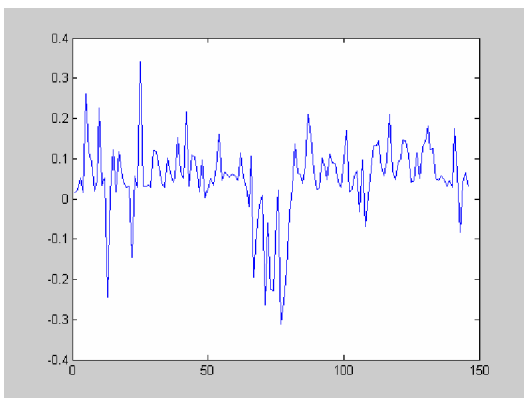


Figure 4 : Voie ACP du signal réel «EMGX»

Cependant, cette méthode pourra être suivie d'une autre étape, permettant de clarifier au mieux les résultats et de minimiser le taux d'erreur. Cette étape sera l'application de l'algorithme de FCM sur le paramètre résultant de la combinaison des différents paramètres du signal.

4.3 - Résultats de la méthode LF :

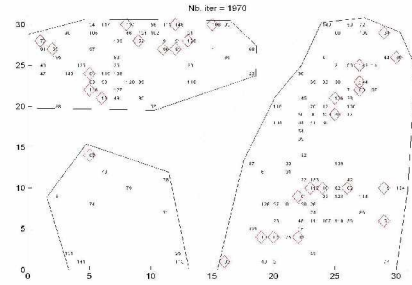


Figure 5 : Méthode LF appliquée sur «EMGX»

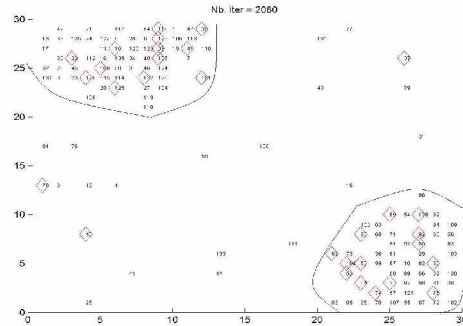


Figure 6 : Méthode LF appliquée sur «EMGY»

Comme le montre les figures 5 et 6 ci-dessus, l'algorithme LF nous donne une première étape pour la classification des EMG. Il permet seulement de les regrouper, suivant leur degré de similarité, en un certain nombre de groupes, avec un certain taux d'incertitude. Pour plus de précision, il a fallu hybrider cette méthode avec une autre, de préférence FCM (paragraphe V). Après ces divers essais, nous avons constaté que le regroupement des éléments dans cette méthode varie d'un essai à l'autre ; et voilà les résultats qu'on a obtenus :

EMGX (taux d'erreur + taux d'incertitude varie entre 30 et 45%)

EMGY (taux d'erreur + taux d'incertitude varie entre 15 et 20%)

5. Comparaison des méthodes

5.1 – Evaluation des différentes techniques :

Bien que le nombre d'échantillons de mesures soit réduit et que les signaux aient été enregistrés sur un seul sujet valide, plusieurs enseignements peuvent être tirés. Il convient de les classer en fonction des critères suivants :

L'interprétabilité : C'est la possibilité d'attribuer une signification aux résultats constitués. En effet, le réseau de neurones travaille comme une boîte noire, il ne donne pas une signification physique liée aux paramètres d'entrée pour les résultats obtenus. Par contre, les algorithmes k-moyenne, LBG et FCM donnent, à la fois, les classes et les centroïdes. Ces derniers représentent, dans notre cas, des énergies dans des bandes de fréquences étroites, ce qui peut

être considéré comme une interprétation physique. Dans l'algorithme LF, le comportement de la fourmi est basé sur la similarité locale entre les données pour pouvoir, par la suite, les regrouper. L'analyse en composantes principales permet de créer une grandeur sur laquelle la classification devient très facile. Mais la corrélation qui relie les paramètres entre eux n'est pas interprétable.

Le temps d'apprentissage et le nombre des itérations: C'est le temps que prend l'algorithme pour faire la classification. On le mesure selon le nombre des itérations. Le RNC utilise un algorithme d'apprentissage pouvant converger lentement. K-moyenne, LBG et FCM peuvent effectuer un nombre conséquent de calculs de distance avant d'obtenir les prototypes mais pourtant, le nombre d'itérations est très petit (6 itérations en moyenne) par rapport au RNC (25 itérations). Une fourmi artificielle est un agent intelligent qui peut faire trois fonctions, pour cela, LF demande un temps d'apprentissage important pour aboutir à la classification, surtout dans le cas des signaux réels. Ce temps d'apprentissage devient négligeable dans ACP dès qu'on arrive à une direction représentative de tous les paramètres, un simple arbre de décision pourra, par la suite, aider à faire la classification.

Le temps de réponse : Une fois que les algorithmes de classification ont terminé leur apprentissage, le temps de la classification de chaque donnée de test est considéré comme un temps de réponse. Ce temps est très rapide dans tous les algorithmes qu'on a essayés, sauf dans LF qui demande, à chaque entrée nouvelle, une recherche de la bonne position de cette entrée.

La réactivité : La réactivité représente la faculté du système de réagir quand les signaux se modifient au cours du temps (Insertion de nouveaux cas dans la base de données). Dans RNC, le fait d'ajouter de nouvelles données impose un relancement d'une phase d'apprentissage pour s'adapter à cette évolution. Il en est de même pour les autres méthodes. Mais, toutes ces méthodes nécessitent moins de temps du fait qu'elles profitent du classement déjà obtenu.

Le taux d'erreur : Il s'agit du nombre d'éléments mal classés. Ce taux est très important dans LF, moindre dans RNC, LBG et K-MOY et minimum dans FCM (comme le montre le paragraphe IV).

La combinaison avec d'autres méthodes : C'est le pouvoir qu'a un algorithme à se combiner avec une autre méthode. Cette propriété concerne LF qu'on doit le combiner avec une autre méthode pour diviser les différents groupes en deux classes; et ACP qui, après avoir trouvé l'axe représentatif de tous les paramètres, utilise n'importe quelle méthode pour faire la classification, de préférence l'algorithme FCM. Celui-ci nous donne de réponses plus précises avec un pourcentage d'appartenance à chaque classe. Concernant les autres algorithmes, on a pu aboutir à la classification sans aucun besoin à d'autres méthodes.

5.2 - Comparaison entre Statistiques et Data Mining :

Dans le domaine du traitement des signaux, les méthodes statistiques gardent une importance assez considérable, vu les résultats trouvés. La méthode des k-moyennes et celle de LBG semblent être favorables pour ce type de problème. Cependant, puisque la quantité de données à traiter après modélisation des signaux est relativement petite, l'utilisation des méthodes comme les réseaux de neurones compétitifs et l'algorithme des fourmis ne se justifie pas. Par contre, la méthode des centres mobiles flous (FCM) a l'avantage de combiner la logique floue à la statistique, et de donner des informations supplémentaires sur les coefficients d'appartenance à chacune des deux classes. Ceci permet, dans certains cas, d'éviter le chevauchement des classes. Ajoutons aussi que l'utilisation du paramètre m avec une valeur supérieure à 2 (3 ou 4 par exemple) et l'augmentation de la valeur d'appartenance (μ_{ik}) à une classe peuvent diminuer le taux d'erreur. Dans notre cas, on a fixé ce coefficient à 0.35 et cela réduit l'erreur à 25% pour le signal EMGX et à 9% pour le signal EMGY.

Ainsi, l'hybridation de l'algorithme des fourmis LF avec FCM est une technique proposée dans la littérature scientifique. Cette technique permet de réduire le nombre de groupes trouvés par les fourmis, et par suite, de réduire le taux d'incertitude, comme le montre la figure 8 :

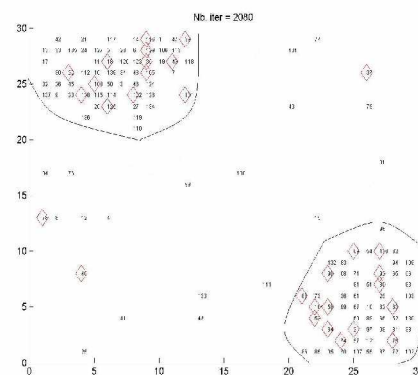


Figure 7 : Signal réel «EMGY» avec LF seul

Groupe G1	Groupe G2
4	3
8	15
12	22
59	31
76	33
77	40
78	41
79	43
84	44
100	123
111	133
131	

Figure 8 : Signal réel «EMGY» avec LF et K-moy

D'autre part, l'hybridation de FCM avec la méthode de l'analyse en composantes principales semble être le meilleur choix pour notre application,

surtout que ACP a réduit le nombre de paramètres et a abouti à la même précision mentionnée dans le paragraphe ci-dessus. Le diagramme suivant (Figure 9.) montre le schéma synoptique proposé pour la classification des signaux EMG :

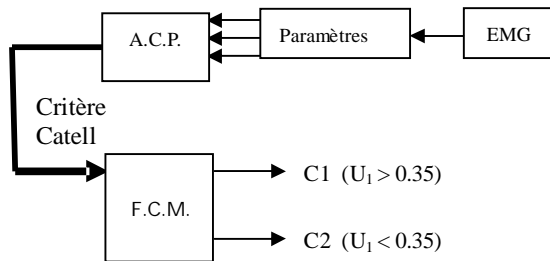


Figure 9: Schéma synoptique de classification des EMG

Conclusion

Ce papier a permis d'initier l'utilisation du processus de Data mining sur des signaux EMG. La sélection des données pertinentes était une tâche essentielle pour aborder correctement le problème de la classification. Un choix judicieux des paramètres a été réalisé : il s'agit d'utiliser les paramètres énergétiques dans des bandes de fréquence différentes.

L'application de plusieurs techniques de classification reconnues en Data mining et en Statistique, constitue une autre étape importante dans ce travail. Nous avons pu, au cours de cet article, pratiquer la méthode des centres mobiles (k-moyennes), la méthode de Linde, Buzo et Gray (LBG), la technique des centres mobiles flous (FCM), l'algorithme des fourmis artificielle (LF), les réseaux de neurones compétitifs (RNC) et l'analyse en composantes principales (ACP). En plus, pour une meilleure séparation de classes, nous avons essayé de mélanger entre deux méthodes (comme ACP+FCM et LF+FCM). Les programmes de ces différentes méthodes ont été réalisés dans l'environnement MatLab.

Les résultats obtenus reflètent précisément le comportement musculaire au cours des trajets de longue durée et permettent de détecter le confort ou l'inconfort des sièges d'un véhicule.

Après une discussion basée sur plusieurs critères et tests d'évaluation, nous avons proposé une hybridation entre ACP et FCM pour traiter notre problème. Surtout que FCM nous permet, par l'ajustement de deux paramètres (μ_{ik} et m), d'avoir deux classes bien séparées.

Pour améliorer notre étude de classification, proposer en perspectives les tâches suivantes : faire un nombre plus important d'acquisition de signaux EMG et y tester les différentes méthodes étudiées ci-dessus; pratiquer d'autres techniques de classification non supervisées, (algorithme génétique, LVQ, GLVQ, GLVQF, PCM, le raisonnement à partir des cas, ...); Essayer de regrouper les segments en trois classes où

cette troisième représentera une phase de transition de l'état de non fatigue à l'état de fatigue; et appliquer le réseau d'ondelettes sur les EMG

Références

- [Dreyfus 2002] Dreyfus G., Martinez M., Gordon M. : *Réseaux de neurones*. Eyrolles. 2002.
- [Dubost 2002] Dubost G. et Tanaka A. "A Wireless NetWork based Biosensor Interface for Music". Proc. Of Intern. Computer Music Conf., sept. 2002.
- [El Falou 2002] El-Falou, "Une approche de la segmentation dans des signaux de longue durée fortement bruités. Application en ergonomie", Thèse de doctorat, UTT, 2002.
- [Huet 2004] Huet S, Glorennec P.Y. "Classification de signaux EMG en vue de la commande automatique d'une main artificielle". Handicap 2004, Paris, 2004.
- [Jorgensen 2000] Jorgensen C., Wheeler K. & Stepniwski S. "Bioelectric Control of a 757 Class High Fidelity Aircraft Simulation". Proceedings of the World Automation Congress, Hawaii, 2000.
- [Khalil 2002] Khalil M., Duchène J. "Uterine EMG Analyzing: A dynamic approach for change detection and classification". IEEE Transaction on Biomedical Engineering, 2002.
- [Knaklitz 2002] Knaflitz M. et Bonato P. "Time-frequency methods applied to muscle fatigue assessment during dynamic contractions". Journal of Electromyography and Kinesiology, vol.9, 2002, pages 337- 350.
- [Linde 1980] Linde Y., Buzo A., Gray R.M. "An algorithm for vector quantization (QV) design". IEEE transactions on communication. 1980.
- [Lumer 1994] Lumer E., Faieta B., "Diversity and Adaptation in Populations of Clustering Ants", Cliff et al. (CLI 94), p. 501-508.
- [Monmarché 2000] Monmarché N., "Algorithmes de fourmis artificielles: applications à la classification et à l'optimisation". Thèse de doctorat, Université de Tours, 2000.
- [Saporta 1999] Saporta G "Probabilités, Analyse des données et Statistiques". Editions Technip, 1999.
- [Younès 2005] M. Atieh Younes & al, "Utilisation des signaux EMG pour la commande intelligente de prothèses de main et de coude", Congrès CITE, Rennes, 2005.