

# Sélection de segments pertinents dans des enregistrements SEMG en conditions expérimentales perturbées

G. FARAH<sup>1,2</sup>, D. HEWSON<sup>1</sup>, J. DUCHÊNE<sup>1</sup>

<sup>1</sup>Université de Technologie de Troyes, ISTIT CNRS, 12 rue Marie Curie, 10010 Troyes Cedex

<sup>2</sup>Renault Technocentre, 1 avenue du Golf, 78288 Guyancourt Cedex

{georges.farah, david.hewson, jacques.duchene}@utt.fr, georges.farah@renault.com

**Résumé** – L’objet de cet article est de proposer une méthode d’extraction de segments pertinents dans des enregistrements électromyographiques de surface (SEMG) en conditions expérimentales perturbées. Nous montrons l’intérêt de l’algorithme CEM pour segmenter les SEMG en classes de bruit de grande amplitude, bouffées d’activité et activité posturale de base. Nous utilisons la modélisation AR afin de classifier les activités EMG exploitables. Un taux de succès de 96% montre l’efficacité de la méthode pour des signaux SEMG enregistrés sur des passagers soumis à des accélérations latérales.

**Abstract** – *The aim of this paper is to extract relevant activities from surface electromyography recordings (SEMG) under difficult experimental conditions with a small signal to noise ratio. We demonstrate that the CEM algorithm is able to segment SEMG signals into high amplitude artifacts, SEMG activity and background postural activity classes. SEMG activities are then identified using the AR modeling. This method produces 96% detection when applied to SEMG recorded on passengers subjected to lateral accelerations.*

## 1. Introduction

Dans le monde automobile, le confort est devenu un critère de qualité qui intéresse tous les constructeurs afin de satisfaire aux demandes et aux besoins de leurs clients. Dans ce contexte, le confort concerne non seulement la qualité d’amortissement ou la ventilation, mais aussi la réduction de la fatigue induite [1] ou la qualité du maintien latéral des sièges. Ce dernier type de confort peut être évalué en utilisant l’électromyographie de surface (SEMG). Même si le SEMG est considéré comme un outil d’évaluation objective du confort, il existe certains problèmes liés à son enregistrement qui rendent difficile le traitement et l’analyse des données.

Les signaux enregistrés sont contaminés par des bruits liés aux conditions expérimentales et présentent souvent un rapport signal sur bruit défavorable. Les enregistrements des activités musculaires posturales sont contaminés par le complexe cardiaque QRS. Les appareils électroniques de mesure génèrent un bruit que nous supposons blanc. Certaines composantes de basses fréquences provenant des mouvements de caisse du véhicule interviennent également dans le signal SEMG. Enfin, des composantes de grande amplitude probablement dues aux contacts des électrodes ou à des jeux de câblage peuvent apparaître lors de l’enregistrement. Ces bruits se manifestent par une très grande variance du signal avec des amplitudes s’approchant de la saturation des amplificateurs de mesure.

Plusieurs méthodes existent pour la détection des complexes QRS. La plupart sont fondées sur les dérivées première ou/et seconde du signal et un seuillage adapté [2]. Le bruit blanc est souvent éliminé par estimation de son niveau sur les hautes fréquences du spectre de puissance ou par la technique des ondelettes [3]. Le bruit basses fréquences quant à lui s’élimine par filtrage passe-haut. L’élimination des bruits de grande amplitude pourra être considérée comme un problème de segmentation.

Le problème lié à ce type de traitement est l’absence de connaissance a priori sur les différentes classes recherchées. Les algorithmes utilisés sont souvent fondés sur des estimations locales et des tests de type GLR [4]. Ces algorithmes présentent par contre un défaut lié au temps de calcul, et aux besoins en ressources mémoire. Une méthode de classification automatique, l’algorithme EM (Expectation-Maximization) a montré ses capacités à séparer les classes dans un modèle de mélange [5]. L’algorithme EM est une technique itérative de maximisation de la loi de vraisemblance en présence de données incomplètes. L’algorithme CEM [5] (Classification Expectation-Maximization) est une version classifiante et rapide de EM. Intuitivement, CEM paraît intéressant pour la séparation entre bruits de grande amplitude, activités musculaires EMG et activités posturales de base, chacune de ces trois classes étant définie par sa variance. Une fois le signal segmenté, les activités EMG exploitables seront identifiées à partir de leur contenu spectral.

L’objectif de cette étude est de proposer une méthode de sélection de segments pertinents dans des enregistrements électromyographiques de surface en conditions expérimentales perturbées. Nous montrons l’apport de l’algorithme CEM dans la séparation des classes. Nous utilisons ensuite la modélisation AR afin de classifier les EMG exploitables. Enfin, la méthode est validée sur des signaux SEMG enregistrés sur des passagers soumis à des accélérations latérales.

## 2. Méthodes

### 2.1 Recueil de données

Dans le cadre d’une étude sur le maintien latéral et la perception de roulis, des essais sur piste ont été réalisés afin

de comparer différents types de véhicules. 18 sujets masculins ont participé à l'étude, avec les caractéristiques anthropométriques suivantes : âge  $43 \pm 8$ , taille  $180 \pm 7$  et poids  $83 \pm 14$ . Les sujets, en tant que passagers, faisaient un tour de piste à vitesse constante (110 km/h), la piste étant choisie afin de refléter une conduite sur route nationale avec virages. 40% du temps correspond à des virages à gauche, 20% à des virages à droite et 40% à une conduite en ligne droite. Les signaux SEMG posturaux ont été enregistrés à l'aide d'électrodes de surface Red Dot TM 2330 pour le niveau dorsal, et Red Dot TM 2238 au niveau de la nuque, de l'abdomen et des cuisses. Le signal cardiaque ECG a été enregistré en même temps pour tous les sujets au niveau du sternum.

L'acquisition des signaux a été réalisée à l'aide d'une carte NI DAQ 6036E connectée sur un PC portable autonome afin de supprimer les rayonnements électromagnétiques pouvant provenir du secteur. Les signaux ont été échantillonnés à 2000 Hz, amplifiés par un facteur 5000, et filtrés passe bande 20-500Hz pour l'EMG et 2-100Hz pour l'ECG. Sous ces conditions expérimentales, les bruits se superposant à l'EMG étaient le complexe QRS, le bruit blanc, les bruits de basses fréquences et les bruits de grande amplitude.

## 2.2 Principe de l'algorithme CEM

L'algorithme EM (Expectation-Maximization) est une technique itérative de maximisation de la loi de vraisemblance en présence de données incomplètes. On l'attribue généralement à Dempster, Laird et Rubin [6]. Nous expliquons cet algorithme sur un modèle de mélange, c'est-à-dire un ensemble de classes chacune définie par ses paramètres.

Soit un échantillon  $X = \{x_1, x_2, \dots, x_n\}$  pour lequel chacune des réalisations  $x_i$  appartient à une classe de l'ensemble  $\{C_k, k = 1 \dots m\}$ . Chacune des classes  $C_k$  est définie par son vecteur de paramètres  $\Theta_k$  et sa probabilité a priori  $\pi_k$ . Le problème consiste alors à estimer le vecteur paramètres  $\Theta = [\pi_1, \pi_2, \dots, \pi_m, \Theta_1, \Theta_2, \dots, \Theta_m]$  par maximisation de la loi de vraisemblance. L'algorithme sera : (a) Solution initiale de paramètres  $\Theta$ , (b) Calcul des probabilités a posteriori, (c) Calcul des nouveaux paramètres qui maximisent la vraisemblance, (d) retour à (b) jusqu'à la convergence.

La difficulté réside dans la maximisation de la log-vraisemblance

$$\log(L(\Theta/X)) = \sum_{i=1}^n \log \left( \sum_{j=1}^m \pi_j f(x_i/\Theta_j) \right)$$

avec  $f(x_i/\Theta_j)$  la probabilité que  $x_i$  provienne de la classe  $C_j$ . L'idée est d'introduire alors  $n$  données manquantes  $Y = \{y_1, y_2, \dots, y_n\}$  avec  $(y_i = k)$  si la réalisation  $x_i$  est générée par la classe  $C_k$  ( $Y$  et  $X$  sont de même dimension  $n$ ). La maximisation de la log-vraisemblance  $\log(L(\Theta/X, Y))$  devient alors envisageable.

L'algorithme CEM est une version classifiante de l'EM, et consiste à ajouter une étape de classification à chaque itération. Les probabilités a posteriori dans ce cas seront égales à 0 ou 1 selon la règle du maximum a posteriori (MAP). Un des grands avantages de CEM par rapport à EM est sa rapidité, ce qui fait qu'il est généralement utilisé dans le cas où des contraintes de temps sont imposées, pour un

fonctionnement en temps réel par exemple. Pour plus de précisions sur EM/CEM, se référer à [5].

L'algorithme CEM appliqué aux modèles de mélange sera dans le cas général :

(a) *Initialisation* :  $\Theta^0 = [\pi_1^0, \pi_2^0, \dots, \pi_m^0, \Theta_1^0, \Theta_2^0, \dots, \Theta_m^0]$

(b) *Estimation* : calcul des probabilités a posteriori

$$\hat{z}_{ik}^{(t)} = \frac{\pi_k^{(t)} f(x_i; \Theta_k^{(t)})}{\sum_{l=1}^m \pi_l^{(t)} f(x_i; \Theta_l^{(t)})}$$

(c) *Classification* : Affectation selon le critère MAP :

$$u_{ik}^{(t)} = \begin{cases} 1 & \text{si } z_{ik}^{(t)} = \max_l(z_{il}^{(t)}) \\ 0 & \text{sinon} \end{cases}$$

(d) *Maximisation* : Maximisation de la vraisemblance

$$\text{classifiante} : \Theta^{(t+1)} = \arg \max_{\Theta} \sum_{i=1}^n \sum_{k=1}^m u_{ik}^{(t)} \log(\pi_k^{(t)} f(x_i; \Theta_k^{(t)}))$$

Revenir à (b) jusqu'à la stabilité de la partition  $U$ .

Le problème de l'algorithme CEM est sa forte dépendance aux conditions initiales. Afin de pallier ce problème, des versions stochastiques de l'algorithme ont été proposées [7]. Une autre possibilité est d'estimer les paramètres de façon non précise, et de les optimiser par CEM.

## 2.3 Séquence de traitement

Dans ce paragraphe, nous décrivons la démarche proposée pour un signal SEMG afin d'en extraire les segments pertinents d'activité musculaire. Au début, nous procédons à une élimination des complexes QRS, puis nous détectons les bruits de grande amplitude avant d'éliminer le bruit blanc et les bruits basses fréquences. Une étape de classification est nécessaire afin de détecter les activités exploitables correspondant à un spectre EMG. Dans la suite, nous désignons par GA les bruits de grande amplitude, PA les activités phasiques et BA l'activité posturale de base.

### 2.3.1 Élimination des complexes QRS

Un signal ECG filtré à bande étroite entre 10 et 20 Hz conserve la forme des complexes QRS. La distribution du signal filtré est gaussienne dans laquelle les pics des QRS apparaissent comme valeurs extrêmes. Un seuil fixé à deux écarts types est suffisant pour les détecter. Le signal est mis à zéro aux alentours des pics, et sur une durée approximative de 100ms, ce qui correspond à la durée du QRS [8].

### 2.3.2 Détection des grandes amplitudes

La première étape consiste à décider si un bruit de grande amplitude existe dans le signal. Le rapport des variances entre un GA et le reste du signal est bien supérieur à celui constaté entre les PA et le BA. De ce fait, la présence de GA implique un minimum dans l'histogramme du logarithme d'énergie (FIG.1). Cette condition n'est pas suffisante car un mélange des classes PA et BA avec un SNR favorable présente aussi un minimum. Pour lever cette ambiguïté, il suffit en plus de tester si la valeur est supérieure à un certain seuil (ici le seuil de linéarité d'amplification). Du coup, un signal comporte une classe GA s'il dépasse le seuil de

linéarité et si un minimum dans l'histogramme de l'énergie existe.

Dans le cas où la classe GA existe, il semble raisonnable d'appliquer CEM deux fois, la première fois pour détecter les GA et la deuxième fois pour détecter les PA. Les lois des trois classes à trouver sont supposées gaussiennes, une condition largement satisfaite pour les PA et les BA. L'algorithme CEM aura alors à trouver trois classes gaussiennes de moyennes nulles et de variances inconnues. Toutes les distributions d'amplitude étant centrées autour de zéro, nous avons choisi de segmenter les signaux à partir de leur distribution d'énergie. Les classes dans ce cas suivent des lois de chi2 centrées. En effet, si  $x_i \sim N(0, \sigma^2)$ , alors

$y = \sum_{i=1}^n x_i^2$  suit une loi de chi2 de densité

$$f(y) = \frac{1}{(2\sigma^2)^{\frac{n}{2}} \Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}}$$

Chacune des classes d'énergie est définie par sa variance  $\sigma^2$ .

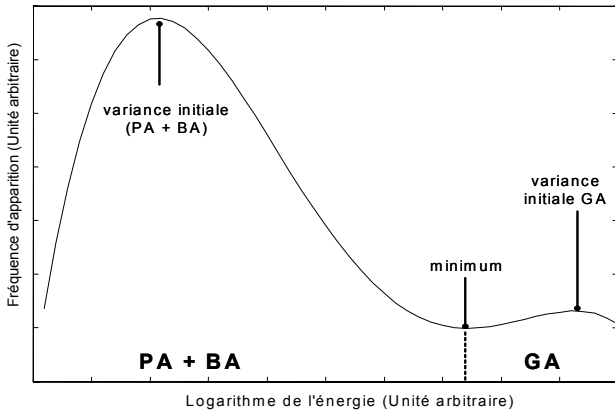


FIG. 1 : Histogramme du logarithme d'énergie

L'algorithme complet pour trouver 2 classes de loi chi2 sera:

(a) Estimation des énergies du signal sur des fenêtres de  $n$  points, ce qui donne un mélange de  $N$  points  $y = \sum_{i=1}^n x_i^2$

(b) Initialisation :  $\Theta^0 = [\pi_1^0, \pi_2^0, \sigma_1^0, \sigma_2^0]$  vecteur initial de paramètres,  $\{k = 1, 2 \text{ et } j = 1 \dots N\}$

(c) Estimation :  $\hat{z}_{jk}^{(t)} = \frac{\pi_k^{(t)} f(y_j; \sigma_k^{(t)})}{\sum_{l=1}^2 \pi_l^{(t)} f(y_j; \sigma_l^{(t)})}$

(d) Classification :  $u_{jk}^{(t)} = \begin{cases} 1 & \text{si } z_{jk}^{(t)} = \max_l(z_{jl}^{(t)}) \\ 0 & \text{sinon} \end{cases}$

(e) Maximisation :  $\sigma_k^2 = \frac{\sum_j y_j u_{jk}^{(t)}}{N \sum_j u_{jk}^{(t)}}$  et  $\pi_k = \frac{1}{N} \sum_j u_{jk}^{(t)}$

Revenir à (c) Jusqu'à la stabilité de la partition  $U$ .

Quand l'algorithme CEM est appliqué à la séparation des classes GA et PA, afin d'accéder la convergence, les

variances initiales sont estimées à partir de l'histogramme des énergies quand un minimum existe (FIG.1). Dans le cas contraire, nous affectons la classe PA à l'ensemble du mélange.

Concernant la séparation des classes PA et BA, 99% du signal d'énergie sont conservés afin d'éliminer les valeurs extrêmes non éliminées par CEM (si GA existait) ou lissées par estimation d'énergie (si GA n'existait pas). Les conditions initiales pour CEM seront définies en fonction des caractéristiques expérimentales.

### 2.3.3 Elimination du bruit blanc et des bruits de basses fréquences

Une décomposition en ondelettes est utilisée afin d'éliminer le bruit blanc [3]. Après une décomposition en 8 niveaux à l'aide des ondelettes de Daubechies5, le signal est reconstruit en éliminant les deux derniers niveaux de détails et en définissant un seuil selon la méthode SURE sur les autres niveaux.

### 2.3.4 Classification

Une fois le signal segmenté, une modélisation AR est utilisée afin de détecter les segments EMG exploitables. Nous nous fondons sur le fait qu'un spectre EMG admet un pôle de grand module pour une phase correspondant à une fréquence aux alentours de 50Hz. Un EMG exploitable appartenant à la classe PA correspond à une activité musculaire phasique. Un segment EMG non exploitable est rejeté.

## 3. Résultats

Les complexes QRS sont tout d'abord éliminés dans l'ECG, et les parties correspondantes dans les signaux SEMG sont mises à zéro (enregistrements synchronisés). Nous avons appliqué l'algorithme CEM une première fois pour détecter les GA avec une fenêtre d'estimation d'énergie de 0,15 seconde, durée obtenue à l'aide d'une courbe COR, correspondant à un taux de 98,7% de bonne détection et un taux de fausse alarme de 8%. Après élimination des valeurs extrêmes du signal d'énergie dépourvu des GA (1%), Nous avons appliqué CEM une deuxième fois avec des fenêtres d'estimation d'énergie de 0,5 seconde correspondant à la durée limite de stationnarité (après application de run test [9]). Dans ce dernier cas, le choix des conditions initiales doit tenir compte des conditions de parcours de la piste : étant donné que 40% de la piste correspond à une conduite sans virage, nous avons estimé la variance des BA par le 20<sup>ème</sup> percentile de l'énergie estimée, et celle des PA par le 80<sup>ème</sup> percentile, ce qui correspond au pourcentage théorique d'activité des muscles sensibles aux virages à droite. Nous avons ensuite appliqué les ondelettes de Daubechies5 pour éliminer le bruit blanc ainsi que les bruits basses fréquences. La figure 2 illustre un exemple de décomposition d'un signal en trois classes : les bruits de grande amplitude, les activités phasiques et les activités posturales de base. Le test de sélection des segments exploitables (signaux de durée 0,5 seconde) était fondé sur la position de la coordonnée (module - phase) du pôle de plus grand module pour un modèle AR d'ordre 20. Nous avons choisi un ordre élevé afin d'obtenir une bonne approximation d'un spectre EMG [10]. Cent segments de la classe PA, aléatoirement choisis, ont été

expertisés comme EMG exploitables ou bouffées de bruit. En parallèle, les coordonnées du pôle de plus grand module ont été calculées. Nous avons remarqué qu'un domaine étroit à l'intérieur du cercle unité, limité en module ( $>0.7$ ) et en phase (fréquences entre 30 et 80 Hz selon les muscles) correspond à des segments EMG exploitables. A l'extérieur, le segment correspond à du bruit. L'application de cette méthode sur un échantillon de 100 segments, choisis aléatoirement dans la classe PA donne un pourcentage de bonne détection de 96% pour un taux de fausse alarme de 9%.

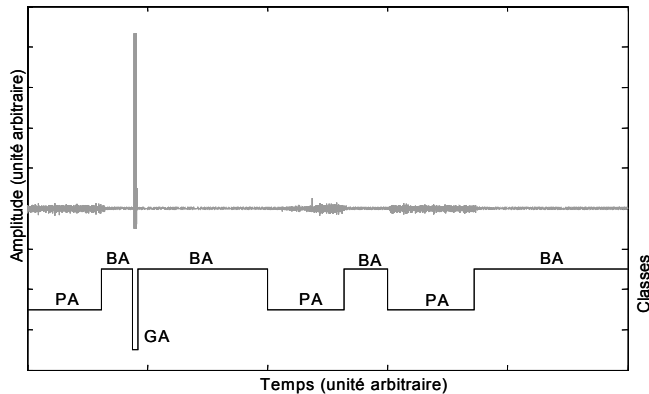


FIG. 2 : Histogramme du logarithme d'énergie

## 4. Discussion

La méthode proposée a montré son efficacité pour l'extraction de segments d'activité musculaire exploitables enregistrés lors de conditions expérimentales perturbées. L'algorithme CEM semble adapté à ce genre de traitement du fait qu'il permet de segmenter les signaux enregistrés sans connaissance a priori des variances de chacune des classes. Cette segmentation n'est pourtant pas aveugle du fait de l'hypothèse d'une distribution gaussienne des amplitudes de chacune des classes. La statistique utilisée est dérivée de l'énergie du signal, correspondant à une loi de  $\chi^2$ . Des statistiques dérivées de l'amplitude ne pouvaient pas être utilisées à cause du passage à zéro de toutes les classes, et donc de leur chevauchement.

La fenêtre de 0,15 seconde pour la détection des GA donne un pourcentage de détection très satisfaisant de 98,7% contre un taux de fausse alarme de 8%, résultat tout à fait acceptable. L'algorithme CEM étant sensible aux conditions initiales, nous nous sommes fondés sur les conditions expérimentales pour définir des conditions initiales aussi proches que possible de la valeur théorique. L'application du CEM deux fois a permis de séparer les activités phasiques des activités posturales de base, cette opération pouvant être utile dans des applications traitant des réponses phasiques d'activité musculaire.

Après segmentation, nous avons classifié les segments EMG exploitables en nous fondant sur la modélisation AR et sur la valeur de la phase du pôle de plus grand module. Bien que cette méthode ait donné de bons résultats, elle a conduit à un taux de fausses alarmes de 9%.

Les méthodes déjà présentées peuvent avoir différentes applications : tout problème de segmentation avec des classes séparables - initialement ou après transformation (i.e.

énergie) - peut être traité à l'aide du CEM. La méthode de classification AR était choisie pour son adaptation à l'EMG, et d'autres méthodes sont envisageables selon les applications.

## 5. Conclusion

Une méthode d'extraction des segments pertinents d'activité musculaire dans des enregistrements en conditions expérimentales perturbées a été proposée. Cette méthode montre l'efficacité de l'algorithme CEM pour la segmentation/classification. Nous avons utilisé des lois de  $\chi^2$  pour chacune des classes et défini les conditions initiales en nous fondant sur les conditions expérimentales. La méthode semble satisfaisante avec 98,7% comme taux de bonne détection pour un taux de fausse alarme de 8%.

La modélisation AR a montré de plus sa capacité à détecter les segments EMG exploitables. Cette méthode semble aussi satisfaisante avec 96% de bonne détection pour un taux de fausse alarme de 9%.

## Remerciements

Les auteurs tiennent à remercier la société automobile Renault pour son support dans le cadre de cette étude et pour l'assistance technique lors de la campagne d'essais sur piste, application directe de la méthode proposée.

## Références

1. Lamotte, T., et al., *Surface electromyography as a tool to study the head rest comfort in cars*. Ergonomics, 1996. **39**(5): p. 781-96.
2. Novak, D., *Processing of ECG signal using wavelets*, in *Czech Technical University*. 2000: Prague.
3. Carre, P., et al., *Denoising of the uterine EHG by an undecimated wavelet transform*. IEEE Trans Biomed Eng, 1998. **45**(9): p. 1104-13.
4. Basseville, M. and I. Nikiforov, *Detection of Abrupt Changes - Theory and applications*, ed. P. Hall. 1993.
5. Govaert, G., *Analyse des données*, ed. Lavoisier. 2003, Paris.
6. Dempster, A.P., N.M. Laird, and D.B. Rubin, *Maximum Likelihood from Incomplete Data via the EM-Algorithm*. Journal of the Royal Statistical Society, 1977(Series B39): p. 1-38.
7. Celeux, G. and G. Govaert, *A classification EM algorithm for clustering and two stochastic versions*. Computational Statistics & Data Analysis, 1992. **14**: p. 315-332.
8. Haueisen, J., et al., *Selective averaging of QRS complexes in magnetocardiograms*. International Journal of Bioelectromagnetism, 2000. **2**(2).
9. Bendat, J. and A. Piersol, *Random Data: Analysis and Measurement Procedures*, ed. Wiley. 1971.
10. Inbar, G.F. and A.E. Noujaim, *On surface EMG spectral characterization and its application to diagnostic classification*. IEEE Trans Biomed Eng, 1984. **31**(9): p. 597-604.