

UNE APPROCHE DE CLASSIFICATION DES CONTRACTIONS UTERINES BASEE SUR LA THEORIE DES ONDELETTES ET LA STATISTIQUE

M.O. Diab^{1,2}, C. Marque¹, M. Khalil²

¹Université de Technologie de Compiègne, UMR CNRS 6600, BP 20529, 60205 Compiègne Cedex, France

²Université Islamique du Liban, Département Biomédical, B.P 30014, Khaldé, Liban
mohamad.diab@utc.fr

(Received 23 November 2004 - Accepted 20 December 2005)

RÉSUMÉ

On propose dans cet article une étude pour classifier les contractions utérines dans le signal électromyographie (EMG) utérin. Comme le contenu fréquentiel dans la contraction change d'une femme à une autre et durant la grossesse, la décomposition par ondelette est utilisée pour l'extraction des paramètres de chaque contraction et une méthode de classification non supervisée basée sur le test de Fisher est utilisée pour classifier les événements. Une projection par Analyse en Composantes Principales est utilisée pour la visualisation des groupes résultants de la méthode de classification.

Les résultats ont montré que les contractions utérines peuvent être classifiées selon des groupes indépendants selon le continu fréquentiel, et selon le terme de grossesse.

Mots clés: ondelette, EMG utérin, classification

ABSTRACT

The purpose of this study is to classify the uterine contractions in the electromyography (EMG) signal. As the frequency content of the contraction changes from one woman to another and during pregnancy, wavelet decomposition is used to extract the parameters of each contraction and then an unsupervised statistical classification method based on Fisher test is used to classify the events. A principal component analysis projection is then used to evidence the groups resulting from this classification.

Results show that uterine contractions may be classified into independent groups according to their frequency content, and in terms of pregnancy.

Keywords: wavelet, uterine EMG, classification

INTRODUCTION

L'accouchement prématuré demeure un problème de santé publique par ses conséquences médicales, psychologiques sur la mère (séparation mère/enfant), par son retentissement économique (Senat *et al.*, 1999). Malgré les progrès de l'obstétrique et de la

néonatalogie, l'accouchement prématuré constitue la plus grande menace pour le fœtus in utero. Il reste la principale cause de mortalité et de morbidité néonatales (Mclean *et al.*, 1993).

Pour l'Organisation mondiale de la santé (loi de 1977), l'expulsion ou l'extraction complète du corps de la mère d'un produit de conception pesant au moins 500 g qui, après cette séparation respire ou manifeste tout autre signe de vie, « est une naissance prématurée ». Si le poids est inconnu, on retiendra un âge gestionnel d'au moins 22 semaines d'aménorrhée (SA) ou une taille d'au moins 25 cm à la naissance (Dubecq *et al.*, 1980).

En France, une loi de 1993, (loi n° 93-22 du 8 novembre 1993, circulaire du 3 mars 1993 et du 22 juillet 1993) dit que tout nouveau-né vivant et viable né après 22 SA ou pesant plus de 500 g doit être déclaré (Horovitz *et al.*, 1996).

On peut donc définir la prématurité comme toute naissance survenant après 22 semaines, d'un enfant de plus de 500 g et avant la 37^e SA révolue (Terzibahcian *et al.*, 1990). Cette définition prend en compte les progrès de la réanimation néonatale et le fait que les études épidémiologiques sur la prématurité comptabilisent les nouveau-nés de plus de 500 g (Terzibahcian *et al.*, 1990).

La fréquence de l'accouchement prématuré est passée en France de 8,2 % des naissances en 1972 à 5,6 % en 1981 et à 4,8 % en 1988-1989. Elle était de 4,3 % en 1990, de 4,2 % en 1991, d'après le service de statistiques des études et des systèmes d'information (SESI). Cette régression concerne surtout la grande prématurité (26 à 32 SA) et les très faibles poids de naissance (inférieurs à 1 500 g) c'est-à-dire les enfants les plus fragiles (Breart *et al.*, 1991).

Les travaux actuels s'intéressent à la caractérisation de l'évolution de la contractilité utérine spontanée au cours de grossesse à risque d'accouchement prématuré. Le but de notre travail est de fournir des méthodes d'identification et de classification qui permettent de classer les contractions du signal EMG utérin (accouchement prématuré ou non).

On propose dans cet article une méthode basée sur la transformée d'ondelette, pour classer les contractions utérines enregistrées, via EMG abdominal, suivie par une méthode de classification non supervisée basée sur le test de Fisher. Cette méthode permet de classer les contractions selon leurs contenus fréquentiels.

Ce papier sera organisé comme suit: tout d'abord on va commencer par la transformée en ondelette ensuite on décrira la méthode de classification. L'application de cette méthode sur les signaux simulés et réels se trouve dans la troisième partie.

MATÉRIELS ET MÉTHODES

Transformée en ondelette

La distribution fréquentielle est l'élément essentiel qui permet de distinguer et identifier les événements dans le signal *EMG*. En effet, la décomposition du signal en ondelettes permet à l'utilisateur de s'adapter plus au signal selon son contenu fréquentiel pour

extraire les informations utiles. Dans ce cas tout changement de fréquence se traduira par un changement sur un ou plusieurs niveaux d'échelles. La classification sera appliquée sur certains signaux de détails et non sur le signal original.

La transformée en ondelette permet d'analyser efficacement des signaux où se combinent des phénomènes d'échelles très différentes. Les paramètres de translation et de dilatation sont les deux arguments de la transformée en ondelette.

La transformée en ondelettes continues d'un signal $x(t)$ prend la forme :

$$T_x^\psi(a,b) = \int x(t)\psi_{ab}(t)dt \tag{1}$$

Dans la plupart des cas, le modèle de la transformée en ondelette n'est pas applicable en continu. Pour une implantation rapide, il est possible d'imposer une composition dyadique de la forme :

$$\psi_{mn}(t) = 2^{m/2}\psi(2^m t - n) \tag{2}$$

$$d_{mn}(t) = \sum x_n(t)\Psi_{mn}(t) \tag{3}$$

Ces équations produisent la décomposition d'un signal sur les échelles. m indique les échelles, n indique la translation dans le temps.

Chaque signal peut se décomposer en détails et approximations, et la forme de la fonction d'échelle est définie par :

$$\phi_{mn}(t) = 2^{m/2}\phi(2^m t - n) \tag{4}$$

Le filtre associé à la fonction $\phi(t)$ est un filtre passe bas et celui associé à l'ondelette est un filtre passe bande. Les formules de détails et des approximations sont :

$$a_x(n,m) = \int_{-\infty}^{+\infty} x(t)\phi_{nm}(t)dt \text{ (Approximations) } \tag{5}$$

$$d_x(n,m) = \int_{-\infty}^{+\infty} x(t)\psi_{nm}(t)dt \text{ (Détails) } \tag{6}$$

Dans ce cas les signaux peuvent être vus comme des détails à différentes échelles (Mallat, 1989).

La formule de reconstruction, permettant de retrouver $x(t)$ à partir de l'ensemble discret de ses coefficients d'ondelettes, devient :

$$x(t) = \sum_{n,m} T_x^\psi(n,m) \psi_{nm}(t) \tag{7}$$

Et on peut montrer que :

$$\text{Détail}_m(t) = \text{approx}_{m-1}(t) - \text{approx}_m(t) \text{ (voir Figure 1) } \tag{8}$$

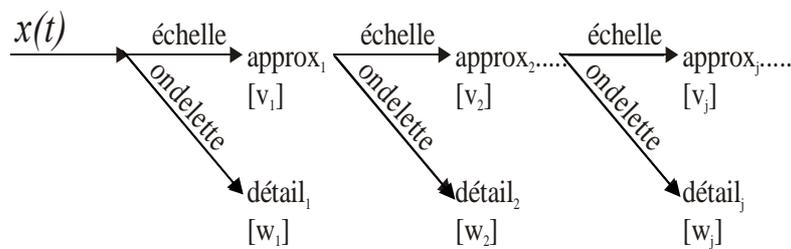


Figure 1. Analyse multirésolution : décomposition successive en approximations et détails.

Choix de l'ondelette

Le choix de l'ondelette analysante est un problème critique dans la théorie d'ondelettes. Pour extraire des événements spécifiques dans un signal, le choix de l'ondelette devient important pour que l'ondelette soit adaptée à la forme de l'événement à détecter.

Dans les domaines d'application très différents, la décomposition en ondelettes est utilisée à des fins de détection des transitoires. Elle a été utilisée dans les domaines de détection des contours des images (Mallat, 2000), la compression (Truchetet, 1998), et dans le traitement de parole, de plus elle a été utilisée pour le débruitage (Marque *et al.*, 1986) et comme outil de classification (Diab *et al.*, septembre 2004 ; december 2004 ; 2005)..

Dans les applications biomédicales, Li (Li *et al.*, 1995) a utilisé une ondelette spline quadratique qui lui permet de distinguer facilement les caractéristiques du bruit et de la ligne de base dans le signal électrocardiographique ECG. C'est une ondelette à support compact et à phase linéaire. Shenhadji (1995) a utilisé une ondelette complexe pour la détection des ruptures dans l'ECG.

Dans cette application, l'ondelette de 'Symlet 5' semble donner une meilleure possibilité de distinction. Des études approfondies sur cette ondelette sont faites dans (Khalil, 1999).

Choix des échelles

Le choix de l'ondelette permet de choisir les niveaux d'échelles les plus appropriés avant d'appliquer le paramétrage utilisant les échelles choisies. Sachant que l'augmentation des valeurs d'échelles donne une composante du signal correspondant à une haute fréquence, donc il faut regarder la bande de fréquence qui correspond à chaque niveau d'échelle afin de choisir les niveaux les plus convenables (Abry, 1997) (Krim & Pesquet, 1995).

Concernant le choix des échelles, nous constatons que nous pouvons nous limiter aux échelles 1 à 5 (Figure 2). Une méthode automatique sur le choix des échelles pour les signaux EMG utérins est appliquée dans (Khalil, 1999).

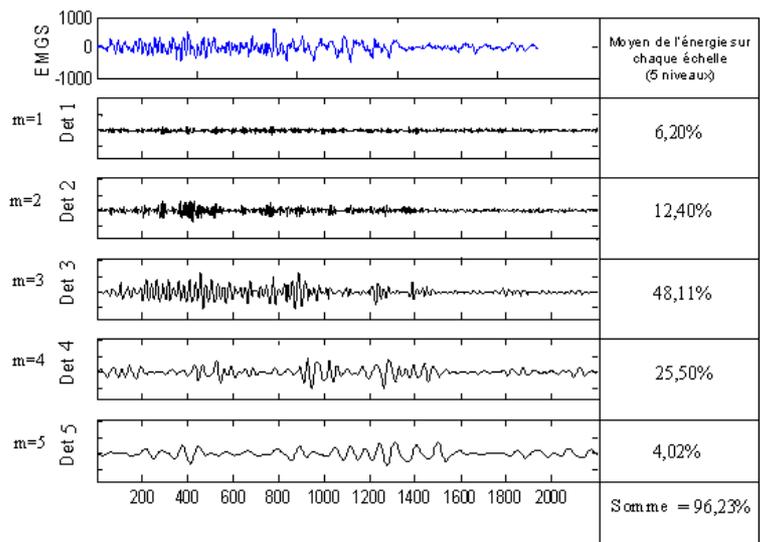


Figure 2. Décomposition par ondelette 'Sym5' sur 5 niveaux d'échelle. La colonne droite montre le pourcentage d'énergie dans les sous signaux de détails. On remarque que 5 sous signaux contiennent 96% de l'énergie totale du signal.

La décomposition en ondelettes est adaptative, elle donne des informations locales sur les évènements. Par contre, la transformée de Fourier donne les caractéristiques fréquentielles globales du signal. C'est pour cela, les ondelettes ont été utilisées pour la détection des évènements dans le signal EMG utérin (Khalil *et al.*, 1998). En plus les énergies

sur les détails d'ondelettes donnent des informations adaptatives et non redondantes, et la somme des détails donne le signal original (ondelettes orthogonales).

Extraction des paramètres

Le paramétrage consiste à calculer les paramètres pertinents qui sont nécessaires pour la classification. Dans l'algorithme (Figure 3), les paramètres utilisés sont les variances sur chaque niveau d'échelle. Donc si on retient d échelles, dans ce cas pour chaque contraction on peut avoir un vecteur de d paramètres. Une fois les paramètres sont calculés, l'algorithme original de classification non supervisée est appliqué.

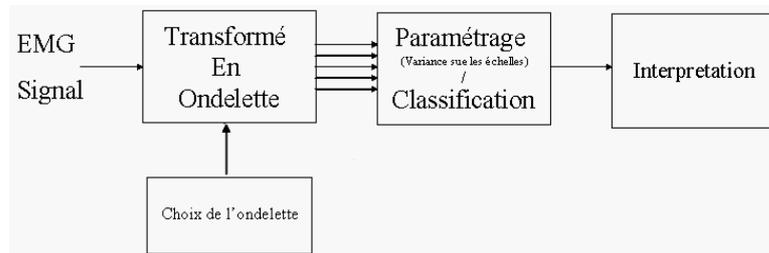


Figure 3. Procédure Paramétrage / Classification: le signal EMG est décomposé en détails, les paramètres sont calculés puis un algorithme de classification non supervisée est ensuite appliqué.

Classification non supervisée

La méthode de classification non supervisée consiste à trouver les vecteurs les plus proches pour les classifier ensemble dans une même classe. Pour cela, après l'extraction des paramètres, les vecteurs sont comparés en utilisant le Test de Fisher mélangé avec la méthode de K-moyenne.

L'algorithme se résume de la façon suivante :

1. En partant du premier vecteur, on l'initialise comme centre de la première classe.
2. Pour chaque nouveau vecteur, calculer la vraisemblance entre ce vecteur et le centre des classes utilisant le test de Fisher (Saporta, 1990) :

$$F_{n_1-1; n_2-2} = \frac{\frac{n_1 S_1^2}{n_1 - 1}}{\frac{n_2 S_2^2}{n_2 - 1}} \quad (9)$$

n_1, n_2 sont les longueurs de deux signaux comparés.

S_1, S_2 sont les variances de deux signaux comparés (signaux des détails décomposés par ondelette).

En fait, on a utilisé le test de Fisher car on suppose que les échantillons du signal suivent une loi normale. Dans ce cas, les variances suivent une loi de Fisher.

1. Si les vecteurs (nouveau vecteur et centre de classe) sont identiques avec une probabilité de confiance donnée, on adapte les classes. Sinon on crée une nouvelle classe.
2. On retourne à l'étape 2.

Un seuil est utilisé pour comparer les vecteurs au centre (intervalle de confiance de Fisher). Un seuil grand aboutit à un nombre minimal de classes. Si la valeur du seuil est petite, alors on obtient un grand nombre des classes.

Après classification, une projection faite par l'Analyse en Composantes Principales (ACP) (Dibuisson, 1995) est utilisée pour visualiser le résultat de la classification. Le but de cette projection est de réduire l'espace des paramètres à un nombre visible par l'utilisateur (2 ou 3).

RESULTATS ET DISCUSSION

La décomposition et la classification sont appliquées initialement sur des signaux simulés. Des segments successifs sont générés en utilisant le filtrage d'un bruit gaussien (car les signaux EMG utérin suivent des lois gaussiennes) avec un filtre passe bande ayant des fréquences centrales différentes mais de même largeur de bande (Figure 4).

Soient H_i (H_j) deux segments correspondant à un filtrage du bruit blanc par un filtre passe bande ayant deux fréquences centrales f_i et f_j .

La résolution de la classification implique l'analyse de la distance entre H_i et H_j i.e.

$$\delta_{ij} = \frac{\Delta F_{ij}}{BP} = \frac{f_i - f_j}{BP}. \tag{10}$$

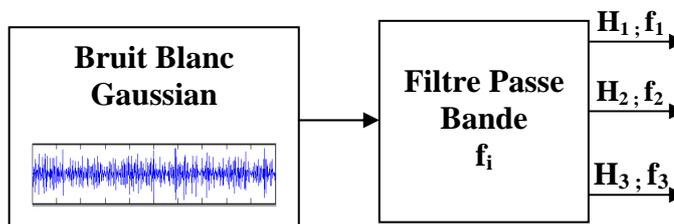


Figure 4. Génération des segments à l'aide des signaux simulés. Cette figure montre la génération des 3 types des segments.

La Figure 5 montre les résultats de classification dans le cas de 2 classes distance de $\delta_{ij} = 0.1$. La Figure 6 montre les mêmes résultats quand on a 3 classes. 20 segments sont générés pour chaque classe ($\delta_{112} = 0.1$, $\delta_{23} = 0.1$, $\delta_{13} = 0.2$) et tous les événements sont bien classifiés.

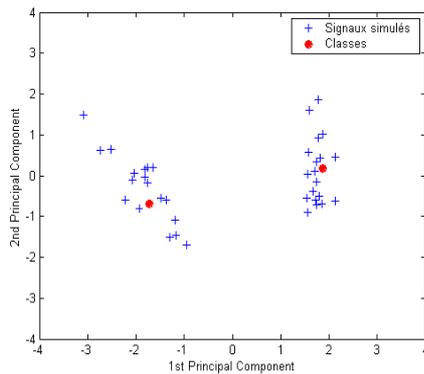


Figure 5. Résultats de classification sur des signaux simulés (2 types de segments ont été générés à l'aide de bruit blanc gaussien).

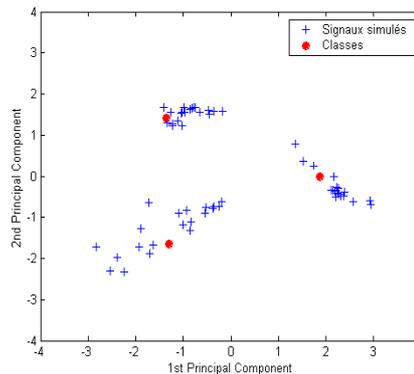


Figure 6. Résultats de classification des signaux simulés. Les segments générés à l'aide de bruit blanc gaussien.

Les signaux réels dont nous disposons ont été enregistrés sur des femmes à plusieurs termes de grossesse et au cours de l'accouchement. Les signaux ont été recueillis à l'aide de deux paires d'électrodes Ag-AgCl de type Beckman. Ces électrodes sont positionnées sur l'abdomen après une préparation soignée de la peau. Le signal électromyographique utérin est recueilli en bipolaire, amplifié et filtré (0.02 – 6 Hz) avant d'être numérisé à une fréquence de 16 Hz et stocké sur une cassette grâce à un enregistreur numérique.

Parmi les signaux recueillis on a appliqué l'algorithme de classification statistique non supervisé sur 10 femmes à 30 SA (semaine d'aménorrhée) et 20 femmes ayant plus de 38 SA.

Les résultats de classification sur les signaux réels sont montrés sur la Figure 7. L'algorithme a donné 2 groupes. Après discussion avec des experts, nous avons remarqué que ces événements ont été classifiés en fonction des semaines d'aménorrhées. En effet, dans cette étape nous avons utilisé 30 femmes (10 femmes à 30 SA et 20 femmes ayant plus de 38 SA). L'algorithme de classification a montré les deux groupes de classes en fonction des semaines d'aménorrhées.

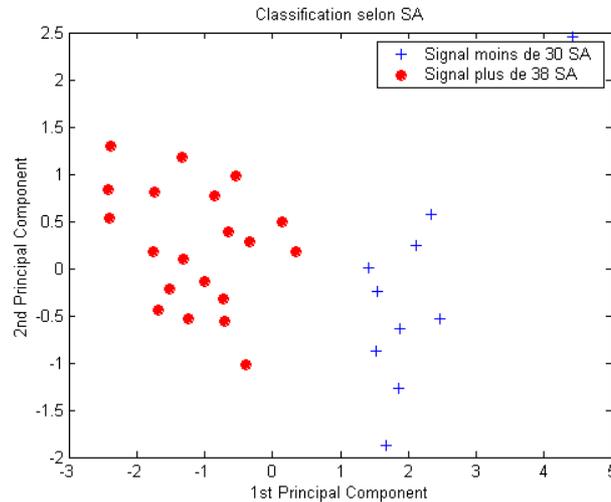


Figure 7. Classification des signaux EMG utérins selon la semaine de grossesse correspondant à l'accouchement ((+) correspond aux femmes qui ont accouché en prématuré, (•) correspond aux femmes qui ont accouché à terme).

Jusqu'à maintenant, on a réussi à classifier les événements par rapport à la semaine de grossesse correspondant à l'accouchement. Ces résultats obtenus permettront de créer un test de classification pour voir si une contraction acquise à une semaine donnée d'une femme test, appartient à un groupe de femmes qui ont accouché en prématuré ou un groupe de femmes qui ont accouché à terme.

CONCLUSION

Une approche de classification en utilisant la décomposition en ondelettes suivie d'une classification statistique a été présentée. D'abord le paramétrage est fait sur les détails en utilisant les variances sur les échelles choisies.

Cette méthode de classification statistique non supervisée est utilisée pour les signaux simulés et ensuite sur les signaux réels. La classification sur des signaux simulés a donné des bons résultats, ce qui valide l'algorithme de classification.

L'application de cet algorithme sur les signaux réels a permis de donner des classes en fonction des semaines d'aménorrhées. 90% des contractions sont bien classifiées.

Le mélange entre décomposition en ondelettes et application d'un test statistique de classification est original. C'est une méthode non supervisée qui peut être adaptée et utilisée pour d'autres applications présentant un changement d'amplitude ou fréquentiel. Il y a un

besoin d'un nombre important des contractions et des femmes pour valider cette théorie. Cette étude doit être continuée afin de comparer cette méthode à d'autres comme celles basées sur les réseaux de neurones et le réseau d'ondelettes.

REFERENCES

- Abry, P. 1997. *Ondelettes et turbulences*, Diderot Editeur. Arts et sciences, Paris.
- Breart, G., Blondel, B., Kaminski, M. 1991. Mortalité et morbidité périnatale en France. *Tournaire Med. Mise à jour en gynécologie obstétrique du Collège National des Gynécologues et Obstétriciens Français*. Paris : Vigot, pp. 175-213.
- Diab, M.O., Marque, C., Khalil, M. 2004. An unsupervised classification method of uterine electromyography signals using wavelet decomposition. *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, San Francisco, CA, USA, September 1-5, pp. 192-195.
- Diab, M.O., Marque, C., Khalil, M. 2004. Uterine electromyography signals: classification and interpretation. *Proceedings of the 2nd Cairo International Biomedical Engineering Conference (CIBEC, IEEE EMBS)*, Cairo, Egypt, December 27-28, pp. 1-4.
- Diab, M.O., Marque, C., Khalil, M. 2005. Unsupervised classification in uterine electromyography signal: toward the detection of preterm birth. *Proceedings of the 27th Annual International Conference of the IEEE EMBS*, Shanghai, China, September 1-4, pp. 2232-2235.
- Dibuisson, B. 1995. *Reconnaissance des formes*, édition HERMES, Paris.
- Dubecq, J.P., Gonnet, J.M., Horovitz, J., Cossard, F., Brachet, M. 1980. Prévention de l'accouchement prématuré par l'association ritodrine-kétoprofène. *10e Journée de la Société Française Méd. Périnat.*, Deauville, pp. 383-385.
- Horovitz, J., Guyon, F., Dubecq, J.P. 1996. Accouchement prématuré. *Éditions Scientifiques et Médicales Elsevier SAS, EMC Consulte*, 5-076-A-10 /1996.
- Khalil, M., Duchêne, J., Marque, C. 1998. Efficacité de la réallocation pour la caractérisation de bouffées EMG dans le plan temps fréquence. *Innov. Tech. Biol. Méd.*, 19(3) : 203-211.
- Khalil, M. 1999. *Une approche de la détection et de la classification dans les signaux non stationnaires. Application à l'EMG utérin*. Thèse de l'Université de Technologie de Troyes.
- Krim, H., Pesquet, J.C. 1995. Multiresolution analysis of a class of non stationary processes. *IEEE Trans. Information Theory*. 41(4): 1011-1020.
- Li, C., Zheng, C., Tai, C. 1995. Detection of ECG characteristic points using wavelet transform. *IEEE transaction, BME.*, 42(1) : 21- 29.
- Mallat, S. 1989. A Theory for multiresolution signal decomposition: the wavelet transform. *IEEE Trans. Pattern Anal. Machine Intell.* 11 : 674-693.
- Mallat, S. juillet 2000. *Une exploration des signaux en ondelettes*. Les éditions de l'Ecole polytechnique, Paris.
- Mclean, M., Walters, A.W.W., Smith, R. 1993. Prediction and early diagnosis of preterm labour: a critical review. *Obstet. Gynecol. Surv.*, 48: 209-25.
- Marque, C., Duchêne, J., Leclercq, S., Panczer, G., Chaumont, J. 1986. Uterine EHG processing for obstetrical monitoring. *IEEE Trans. BME.*, 33 : 1182-1187.
- OMS 1977.
- Saporta, G. 1990. *Probabilités et Analyse de données statistiques*. Edition Technip, Paris.

- Senat, M.V., Tsatsaris, V., Ville, Y. et Fernandez, H. 1999. Menace d'accouchement prématuré. *Encycl. Méd. Chir., (Elsevier, Paris), Urgences*, 24-213-A-20, 17 p.
- Shenhadji, L., Bellanger, J.J., Carraut, G. 1995. Détection temps échelle d'événements paroxystiques intercriptiques en électroencéphalogramme. *Traitement du Signal*, 12(4): 357-371.
- Terzibahcian, J.J., Mida, M., David, E. 1990. Bilan des menaces d'accouchement prématuré : étiologie, conduite à tenir, résultats. *Rev. Fr. Gynécol. Obstét.*, 85 : 673-678.
- Truchetet, F. 1998. *Ondelettes pour le signal numérique*. Collection traitement du signal, édition HERMES, Paris.