

Détection de bouffées dans l'EEG du nouveau-né prématuré

Fabienne PORÉE^{1,2}, Xavier NAVARRO³, Mathieu KUCHENBUCH^{1,2,4}, Alain BEUCHÉE^{1,2,4}, Guy CARRAULT^{1,2,5}

¹Université de Rennes 1, LTSI, Rennes, France

²INSERM, U1099, Rennes, France

³Sorbonne Universités, UPMC Université Paris 06, INSERM UMRS-1158, Paris, France

⁴CHU Rennes, Pôle de pédiatrie médico-chirurgicale et génétique clinique, Rennes, France

⁵INSERM, CIC-1414, Rennes, France

Fabienne.Poree@univ-rennes1.fr, Xavier.Navarro@upmc.fr, Mathieu.Kuchenbuch@chu-rennes.fr,
Alain.Beuchee@chu-rennes.fr, Guy.Carrault@univ-rennes1.fr

Résumé – Ce travail propose et compare deux méthodes de détection des bouffées EEG chez le nouveau-né prématuré : une méthode de seuillage et une méthode, nouvelle dans ce contexte, basée sur la régression logistique. L'évaluation est réalisée sur une base de 10 bébés d'âge gestationnel compris entre 25 et 30 semaines, et d'âge postnatal supérieur à sept semaines. Après avoir établi une référence basée sur les annotations manuelles de trois experts, nous montrons que la régression logistique fournit les meilleures performances et qu'elles sont comparables aux accords inter-experts.

Abstract – This paper proposes and compares two automated detectors of EEG bursts based on different classifiers: a thresholding classical method and a newly applied logistic regression algorithm. Both solutions are evaluated on a database of 10 premature newborns 25 to 30 weeks gestational age, and at least seven weeks postnatal age. After establishing a gold standard from the visual marks of three clinicians, we show that the performances of the logistic regression-based classifier over-perform those obtained by the threshold detector and are comparable to inter-rater agreements.

1 Introduction

Le nouveau-né prématuré naît avec plusieurs de ses fonctions vitales immatures, parmi lesquelles l'activité cérébrale. L'électroencéphalogramme (EEG) est alors caractérisé par des tracés discontinus (ou tracés alternants), alternant des périodes de calme avec des périodes d'ondes lentes et d'amplitude élevée, appelées bouffées ou "bursts", et évoluant avec l'âge. En particulier, il a été montré une diminution des intervalles inter-bouffées (inter-burst intervals, IBI) avec l'âge des enfants [1], rendant cette information pertinente pour l'étude de la maturation des nouveau-nés prématurés.

La détection des bouffées et des IBI, effectuée manuellement par les cliniciens, est particulièrement longue à réaliser mais aussi dépendante de l'expert. Ainsi, des algorithmes ont été proposés dans la littérature afin d'automatiser cette procédure. Différentes approches, basées sur des calculs de seuils, ont été proposées à ce jour, comme le NLEO (nonlinear energy operator) chez l'adulte [2] ou les SVM (support vector machines) chez le nouveau-né à terme [3]. Plus récemment Palmu et al. [4] ont proposé une méthode pour le grand prématuré, d'âge gestationnel (GA) compris entre 23 et 30 semaines, après trois jours de vie, basée sur l'opérateur NLEO.

Ce travail s'adresse à des prématurés d'âge gestationnel compris entre 25 et 30 semaines, ayant au moins sept semaines d'âge postnatal. Deux méthodes de classification sont proposées. La première est basée sur un seuillage, où trois fonctions d'activité (dont l'opérateur NLEO) sont testées et la seconde utilise la régression logistique [5]. Ces deux approches sont évaluées en comparant les résultats de classification à des détections manuelles réalisées par trois experts différents.

2 Méthode

2.1 Comparaison à un seuil

Le premier classifieur (Figure 1) est inspiré de la méthode proposée par Palmu et al. [4], à laquelle plusieurs modifications ont été appliquées. Tout d'abord, un filtrage

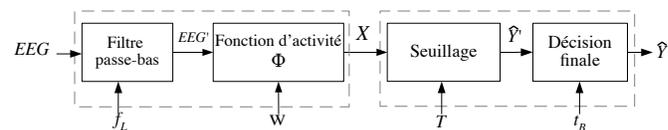


FIGURE 1 – Diagramme de la méthode de seuillage.

passé-bas de fréquence de coupure $f_L=8$ Hz est appliqué, pour supprimer les fréquences inutiles, le tracé alternant étant principalement constitué d'ondes delta et theta.

Ensuite, une fonction d'activité Φ , appliquée dans des fenêtres glissantes de $W=1s$ de longueur $l=Wf_s$, où f_s est la fréquence d'échantillonnage, pouvant prendre trois définitions différentes, est calculée :

1. Mm : La différence entre les valeurs maximales et minimales :

$$\Phi_{Mm} = \max_{k=1,\dots,l}\{EEG(k), EEG(k) \geq 0\} - \min_{k=1,\dots,l}\{EEG(k), EEG(k) < 0\}. \quad (1)$$

2. ADIF (Amplitude Difference measure) : La somme des valeurs absolues divisée par le nombre d'échantillons [6] :

$$\Phi_{ADIF} = \frac{1}{l} \sum_{k=1}^l |EEG(k)|. \quad (2)$$

3. NLEO (Nonlinear energy operator) : nous avons utilisé la version proposée par Särkelä et al. [7] :

$$\Phi_{NLEO} = \frac{1}{l} \sum_{k=1}^l |NLEO(k)|. \quad (3)$$

avec :

$$NLEO(k) = \begin{aligned} & EEG(k)EEG(k-3) \\ & - EEG(k-1)EEG(k-2). \end{aligned} \quad (4)$$

Le vecteur de paramètres X est ensuite comparé à un seuil T . Si $X(i) > T$, un burst est détecté et $\hat{Y}' = 1$, sinon $\hat{Y}' = 0$. Dans le dernier bloc, et comme dans [4], les bursts dont la longueur est inférieure à $t_B=1$ s sont éliminés dans le signal final \hat{Y} .

2.2 Régression logistique

Le deuxième classifieur (Figure 2) est basé sur la méthode de régression logistique [5], particulièrement adaptée à la classification binaire, dont la sortie donne la probabilité de présence de burst (valeur entre 0 et 1). Cette mé-

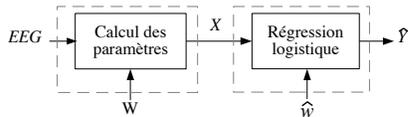


FIGURE 2 – Diagramme de la méthode basée sur la régression logistique.

thode ne nécessite pas de filtrage puisque les paramètres extraits sont des variables qui vont faire l'objet d'une régression statistique, ce qui contribue au débruitage. Neuf paramètres sont calculés dans des fenêtres glissantes de $W=1s$. Ils ont été choisis dans le but de fournir différents types d'information sur les signaux (amplitude, statistique, énergie et fréquence).

1. Mm : Calcul de Mm dans la fenêtre courante.
2. Mm_{-1} : Calcul de Mm dans la fenêtre précédente.
3. Mm_{+1} : Calcul de Mm dans la fenêtre suivante.
4. MD : Maximum de la différence absolue :

$$MD = \max_{k=1,\dots,l} \{|EEG(k) - EEG(k-1)|\}. \quad (5)$$

5. Sd : Ecart-type.
6. Kt : Kurtosis.
7. NL : Calcul du NLEO (éq. (4)).
8. RMS :

$$RMS = \sqrt{\frac{1}{l} \sum_{k=1}^l EEG^2(k)}. \quad (6)$$

9. Dm : Différence moyenne :

$$Dm = \frac{1}{l} \sum_{k=1}^l |EEG(k) - EEG(k-1)|. \quad (7)$$

Enfin, la fonction logistique est appliquée au vecteur de paramètres X avec les coefficients de régression \hat{w} estimés lors de l'étape d'apprentissage, et fournit le vecteur \hat{Y} .

3 Résultats

3.1 Protocole

Un total de $N=10$ nouveau-nés prématurés, d'âge gestationnel compris entre 25 et 30 semaines et ayant au moins sept semaines d'âge postnatal, ont été inclus dans l'étude. Pour chacun d'eux, quatre voies EEG (Fp1, Fp2, T3, T4) ont été enregistrées à la fréquence $F_s=512$ Hz. La détection de bouffées a été appliquée sur des extraits de durée $D=300$ s, choisis en période sommeil calme, pour être sûrs de disposer de tracés alternants, sur la dérivation (Fp1-T3 ou Fp2-T4) de meilleure qualité.

Les EEG ont été sous-échantillonnés à 128 Hz puis débruités pour supprimer les artefacts et le bruit à l'aide d'une méthode que nous avons développée, combinant une décomposition modale empirique et un filtrage adaptatif [8].

Trois experts néo-natologistes (A, B, C) du CHU de Rennes ont ensuite été sollicités pour annoter les EEG, en terme de présence ou absence des bouffées, et ceci à deux reprises. À partir des annotations, ont été définies :

- l'annotation finale de chaque expert, définie comme la zone d'accord (sur les bursts et les IBI) entre ses deux annotations,
- la référence, ou gold standard, définie comme la zone d'accord entre les annotations finales des trois experts (Figure 3).

Ensuite, les pourcentages d'accord (AR) ont été calculés entre les deux annotations d'un même expert, entre les annotations d'experts différents et entre les méthodes et le gold standard, en calculant le ratio entre les valeurs concordantes (burst ou IBI) et la taille de l'échantillon.

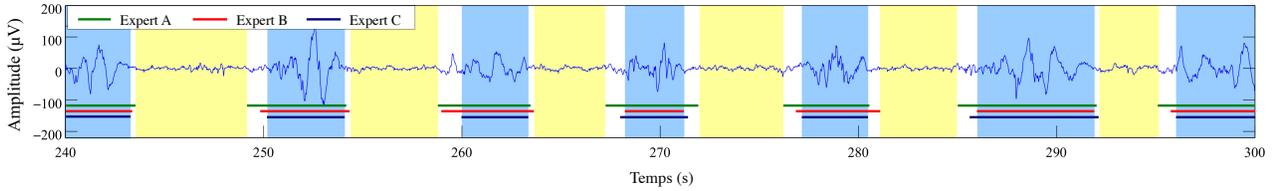


FIGURE 3 – Exemple d’une minute annotée par les trois experts. Les zones de désaccord (en blanc) ne sont pas retenues pour la définition du gold standard.

3.2 Performances des méthodes de détection

Le réglage des paramètres des deux approches (les trois méthodes de seuillage et la méthode LR) a été réalisé suivant une procédure leave-one-out, en incluant neuf bébés pour l’apprentissage, et en réalisant le test sur le bébé restant. La procédure est donc répétée dix fois, en validant à chaque fois sur un bébé différent [9].

Concernant les méthodes de seuillage, cette procédure a permis le réglage du seuil T en maximisant l’accord entre le résultat de chaque méthode et le gold standard. Les valeurs de T obtenues sont $6\mu V$ pour la méthode Mm, $34\mu V$ pour la méthode ADIF et $3.3\mu V$ pour le NLEO. Concernant la régression logistique, le critère d’inclusion des bursts a été fixé à 0.5 puis les coefficients \hat{w} du modèle ont été ajustés.

Les quatre méthodes ont pu être ensuite comparées en termes de sensibilité (Se), spécificité (Sp), accord (AR_Y) et aire sous la courbe COR (AUC). Les valeurs représen-

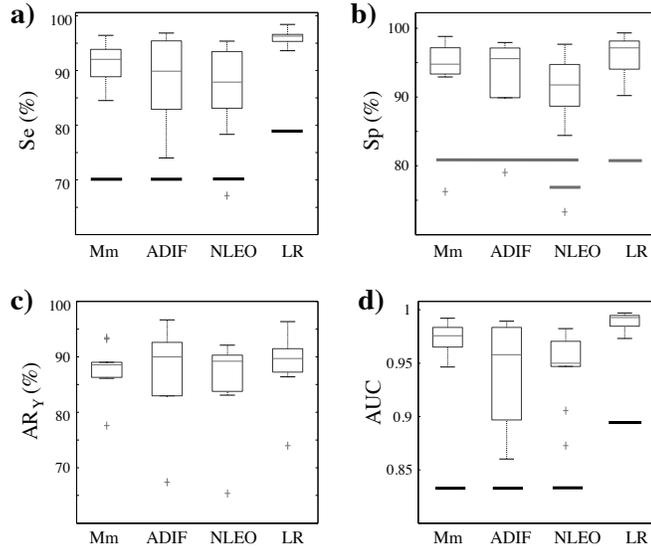


FIGURE 4 – Comparaison de performances obtenues avec les quatre méthodes (Mm, ADIF, NLEO, LR). Les lignes horizontales de hauteurs différentes représentent une différence significative obtenue après un test de Mann-Whitney (noir : $p < 0.01$; gris : $p < 0.05$).

tées sur la Figure 4 montrent la supériorité de la méthode LR dans les quatre cas, avec en plus des différences significatives pour la sensibilité et l’aire sous la courbe COR, après un test de Mann-Whitney.

3.3 Comparaison entre les détections manuelle et automatique

On s’intéresse, dans cette partie, à la comparaison entre les annotations intra- et inter-experts et l’approche automatique. La méthode LR ayant fourni les meilleures performances, seule celle-ci est considérée.

3.3.1 Accords entre experts

La Table 1(A) résume les accords entre experts obtenus sur l’ensemble des données annotées, pour chaque expert entre ses deux annotations, entre deux experts (trois cas) et entre les trois experts. L’accord intra-expert est compris entre 85.7 et 86.1% et l’accord inter-experts entre 76.5 et 81.9% entre deux experts. Il est égal à 78.7% entre les trois experts. Cette valeur, plus faible que dans [10], où 81% et 86% d’accord avait été obtenus chez des bébés respectivement très et extrêmement prématurés. Ceci peut sans doute s’expliquer par le fait que notre étude s’applique à des bébés plus matures, puisque enregistrés à sept semaines d’âge postnatal.

3.3.2 Accord global

L’accord moyen sur les 10 bébés entre les annotations des trois experts et la méthode LR est présenté dans la Table 1(B). Il est compris entre 70.5 et 82.8%. Il est aussi intéressant de noter que deux des trois valeurs (LR-B et LR-C) sont supérieures aux accords inter-experts de la Table 1(A).

3.3.3 Accords sur les caractéristiques

La qualité des détections fournies par la méthode LR est maintenant évaluée de manière qualitative, en analysant : i) le nombre de bursts, ii) la durée moyenne des bursts et iii) le pourcentage de bursts. Comme le montre la Figure 5, les valeurs (moy \pm std) obtenues après la détection automatique (LR) sont du même ordre de grandeur que celles fournies par les annotations des experts.

TABLE 1 – (A) Accords (%) intra- et inter-experts. (B) Accords (%) entre la méthode de régression logistique et les trois experts.

(A)						(B)		
Intra-expert			Inter-experts (2 à 2)			LR-Expert		
AR_A	AR_B	AR_C	AR_{AB}	AR_{AC}	AR_{BC}	AR_{LR-A}	AR_{LR-B}	AR_{LR-C}
85.7	85.9	86.1	76.5	77.8	81.9	70.5	82.8	81.2

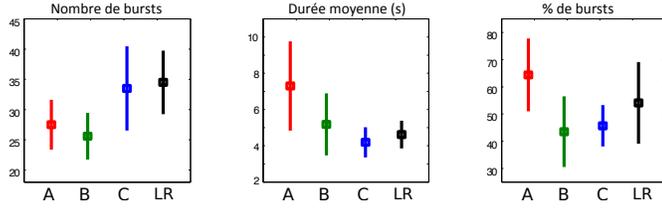


FIGURE 5 – Caractéristiques (moy \pm std) des bursts d’après les experts (A, B, C) et la méthode LR.

Une analyse statistique de type Mann-Whitney a été réalisée sur les caractéristiques des bouffées (Table 2). Des différences significatives sont observées à la fois entre deux experts différents et entre chaque expert et la méthode LR.

TABLE 2 – p -values calculées sur les caractéristiques des bursts entre deux experts différents et entre chaque expert et la méthode LR (en gras $p < 0.05$).

	Nombre	Durée moyenne (s)	% de bursts
A vs. B	0.289	0.045	0.005
A vs. C	0.088	0.002	0.007
B vs. C	0.011	0.241	0.473
LR vs. A	0.010	0.005	0.089
LR vs. B	0.002	0.970	0.241
LR vs. C	0.879	0.212	0.241

4 Discussion et Conclusion

Dans cette étude, nous avons proposé et comparé deux approches pour la détection des bouffées dans l’EEG du prématuré. La seconde approche, basée sur la régression logistique a fournit des résultats supérieurs à ceux obtenus avec l’approche basée sur un seuillage.

Un point fort de ce travail est la réalisation d’un gold standard basé sur les annotations de trois experts différents.

Les résultats obtenus avec la régression logistique ont montré que la détection automatique permet d’obtenir des performances comparables aux accords inter-experts, en terme d’accord global mais aussi de caractéristiques des bouffées.

Références

- [1] J. S. Hahn, H. Monyer, and B. R. Tharp, “Interburst interval measurements in the EEGs of premature infants with normal neurological outcome,” *Electroencephalography and Clinical Neurophysiology*, vol. 73, no. 5, pp. 410–418, Nov. 1989.
- [2] J. Kaiser, “On a simple algorithm to calculate the ‘energy’ of a signal,” in *ICASSP’90*, 1990, pp. 381–384 vol.1.
- [3] J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg *et al.*, “Classification of burst and suppression in the neonatal electroencephalogram,” *Journal of Neural Engineering*, vol. 5, no. 4, p. 402, Dec. 2008.
- [4] K. Palmu, S. Wikström, E. Hippeläinen, G. Boylan *et al.*, “Detection of ‘EEG bursts’ in the early preterm EEG : visual vs. automated detection,” *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1015–1022, Jul. 2010.
- [5] D. W. Hosmer Jr and S. Lemeshow, *Applied logistic regression*. John Wiley & Sons, 2004.
- [6] A. Varri, “Digital processing of the EEG in epilepsy,” Ph.D. dissertation, Tampere University of Technology, Finland, 1988.
- [7] M. Särkelä, S. Mustola, T. Seppänen, M. Koskinen *et al.*, “Automatic analysis and monitoring of burst suppression in anesthesia,” *Journal of Clinical Monitoring and Computing*, vol. 17, no. 2, pp. 125–134, Feb. 2002.
- [8] X. Navarro, F. Porée, and G. Carrault, “ECG removal in preterm EEG combining empirical mode decomposition and adaptive filtering,” in *ICASSP’12*. IEEE, 2012, pp. 661–664.
- [9] X. Navarro, “Analysis of cerebral and respiratory activity in neonatal intensive care units for the assessment of maturation and infection in the early premature infant,” Ph.D. dissertation, Université Rennes 1, 2013.
- [10] K. Palmu, N. Stevenson, S. Wikström, L. Hellström-Westas *et al.*, “Optimization of an NLEO-based algorithm for automated detection of spontaneous activity transients in early preterm EEG,” *Physiological Measurement*, vol. 31, no. 11, pp. N85–93, Nov. 2010.