

# Une approche non supervisée pour l'estimation du sommeil calme chez des nouveau-nés prématurés et à terme

Houda JEBBARI Sandie CABON Patrick PLADYS Guy CARRAULT Fabienne PORÉE

Univ Rennes, CHU Rennes, INSERM, LTSI - UMR 1099, F-35000, France

**Résumé** – Une méthode d'estimation du sommeil calme pour le nouveau-né prématuré et à terme est proposée dans cet article. Elle s'appuie sur la fusion du mouvement, extrait de vidéos, de l'électrocardiogramme et de la respiration. L'originalité de cette approche est d'être auto-apprenante et repose sur l'utilisation d'un algorithme de clustering non supervisé pour l'estimation des périodes de sommeil calme dans les signaux cardio-respiratoires. Les performances sont évaluées sur 25 enregistrements de 8 heures, ce qui conduit à un balanced accuracy moyen de 76%, et de 80% pour les bébés d'âge post-menstruel supérieur à 33 semaines.

**Abstract** – A method for estimating quiet sleep for preterm and full-term newborns is proposed in this article. It is based on the fusion of motion, extracted from videos, electrocardiogram, and respiration. The originality of this approach is that it is self-learning and relies on the use of an unsupervised clustering algorithm for the estimation of periods corresponding to quiet sleep in cardiorespiratory signals. Performances are evaluated on a set of 25 recordings of 8 hours, which leads to an average balanced accuracy of 76%, and 80% for neonates with a post-menstrual age superior to 33 weeks.

## 1 Introduction

Les bébés nés prématurément, c'est-à-dire avant 37 semaines d'Age Gestationnel (AG), sont exposés à divers problèmes de santé parce qu'ils naissent à un moment critique du développement du système nerveux central. Leur état de santé est évalué, en Unités de Soins Intensifs Néonataux (USIN), par la surveillance continue de plusieurs signes vitaux (activité cardiaque, respiration...), mais aussi par l'analyse du sommeil, celui-ci étant directement lié à la maturation du cerveau [4].

Classiquement, l'évaluation du sommeil est accessible par la polysomnographie, mais celle-ci n'est pas beaucoup utilisée chez le prématuré car elle nécessite un nombre important de capteurs (en particulier pour l'électroencéphalogramme), et ne peut pas être appliquée sur de longues durées. On lui préfère donc l'observation comportementale (activité corporelle, état des yeux, rythme cardio-respiratoire, vocalisations...), pour annoter les quatre stades de sommeil du nouveau-né, à savoir le Sommeil Calme (SC), le Sommeil Agité (SA), le Sommeil Indéterminé (SI) et l'Eveil [1]. Cette technique est en revanche très chronophage et le développement de méthodes automatiques d'estimation des stades de veille et de sommeil des nouveau-nés est primordial.

Dans la littérature récente, des méthodes supervisées ont été mises en œuvre à partir des signaux cardiaques et/ou respiratoires. Des résultats encourageants ont été obtenus en utilisant des stratégies de seuillage [6], d'apprentissage classique [10] ou profond [11]. Afin de prendre en compte les aspects comportementaux du sommeil du nouveau-né, d'autres stratégies ont porté sur le traitement de vidéos [2]. Cependant, toutes ces approches rencontrent des difficultés lorsqu'elles sont appliquées séparément, en particulier pour séparer certains stades.

Dans cet article, nous proposons une approche combinant le traitement de vidéos et de signaux électrophysiologiques. On se focalisera sur le sommeil calme, car c'est le stade le plus important en termes de maturation [4]. Le SC étant caractérisé par une absence d'activité motrice et un rythme cardio-respiratoire régulier, la méthode consiste à détecter puis fusionner ces infor-

mations. Pour cela, on s'appuie sur un algorithme de clustering non supervisé permettant de séparer naturellement les classes (SC vs non-SC) dans les signaux cardio-respiratoires. Il s'agit d'une stratégie auto-apprenante spécifique-bébé, permettant de répondre à la diversité de la pratique clinique en USIN. Par ailleurs, plusieurs stratégies de fusion sont comparées.

L'article est organisé comme suit. La section 2 décrit le protocole, la base de données et les méthodes mises en œuvre. Dans la section 3, les résultats obtenus sont présentés. La discussion et la conclusion sont données dans la section 4.

## 2 Méthodes

Dans cette section, nous décrivons d'abord le protocole d'acquisition et la base de données. Ensuite, la méthodologie est détaillée en trois parties : (i) la détection des intervalles de non-mouvement, (ii) la détection des segments réguliers dans l'ECG et dans la respiration (iii) la fusion des trois modalités pour l'estimation du SC (voir Figure 1).

### 2.1 Protocole et base de données

La base de données de cette étude reprend une partie de la base de données du projet européen Digi-NewB [9]. Elle contient des vidéos enregistrées avec des caméras infrarouges et des signaux physiologiques. Les vidéos ont été acquises à la fréquence de 25 images par secondes avec un encodage MPEG-4. Des électrodes permettent d'enregistrer la respiration (par impédancemétrie) et l'ECG à une fréquence d'échantillonnage de 62.5 et 500 Hz, respectivement. Cette étude comprend 15 nouveau-nés, dont aucun ne présentait de développement pathologique selon l'examen médical. La population est divisée en trois groupes en fonction de l'AG (en semaines) des bébés :

- Groupe 1 : 5 grands prématurés ( $27 \leq AG < 29$ );
- Groupe 2 : 5 moyens prématurés ( $33 \leq AG < 36$ );
- Groupe 3 : 5 nés à terme ( $AG \geq 39$ ).

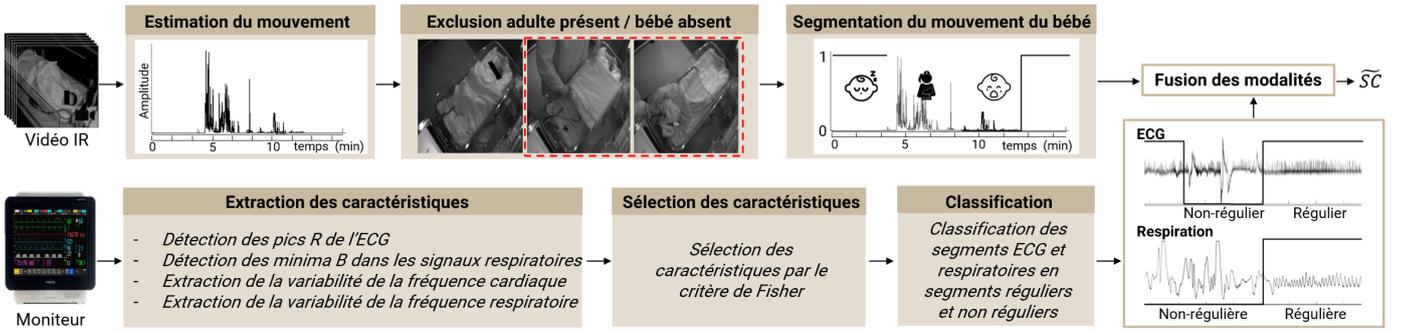


FIGURE 1 : Principe général de la stratégie d'estimation du SC.

Tous les nouveau-nés ont été enregistrés au cours de la première semaine de vie (jour 1) et, pour les groupes 1 et 2, ils ont également été enregistrés à l'approche de leur sortie de l'hôpital (jour 2), à environ 37 semaines d'Age Post-Menstruel (APM), qui correspond à l'AG plus l'âge chronologique. Les enregistrements ont été effectués la nuit entre 22 heures et 6 heures. Cela a donné lieu à 25 enregistrements (200 heures).

Pour chaque enregistrement, les phases de sommeil calme ont été annotées manuellement par un expert [1, 4]. Des périodes dites "NOK" ont également été annotées. Elles désignent les périodes non analysables car le bébé n'est pas dans le lit ou en l'absence des signaux physiologiques. Les périodes restantes ont été automatiquement considérées comme les états non-SC (NSC), et comprennent les états SA, SI et Eveil.

## 2.2 Traitement du mouvement

La méthode d'extraction du mouvement repose sur le traitement des vidéos (voir [3] pour plus de détails) et s'effectue en trois étapes (Figure 1). Tout d'abord, une différence inter-images est calculée pour estimer le mouvement. Ensuite, les périodes comprenant l'absence du bébé ou la présence d'un adulte sont identifiées grâce à un modèle par apprentissage profond. Enfin, une segmentation du mouvement est réalisée pour estimer les intervalles de mouvement et de non-mouvement ("NoM") du bébé, en utilisant un classifieur de type Random Forest. La sortie est un signal égal à "0" pendant les phases de mouvement, à "1" pendant les phases de non-mouvement et à "NaN" en cas d'absence du bébé ou de présence d'adultes.

## 2.3 Traitement de l'ECG et de la respiration

### 2.3.1 Extraction des caractéristiques

Les pics R du signal ECG sont détectés à l'aide d'une version modifiée de l'algorithme de Pan et Tompkins, avec des coefficients de filtrage spécifiquement adaptés aux nouveau-nés [5]. Les débuts de phase d'inspiration B dans les signaux respiratoires sont détectés à l'aide d'un algorithme développé dans [7]. Étant donné que la durée minimale d'un intervalle de SC est fixée à 20 secondes selon les annotations, les pics R et les débuts de phase d'inspiration B sont segmentés par une fenêtre glissante de 40 secondes avec un chevauchement de 50%. La librairie Neurokit2 Python est utilisée pour transformer ces séries temporelles en caractéristiques de Variabilité de la Fréquence Cardiaque (VFC) et de Variabilité de la Fréquence Respiratoire (VFR) [8]. Au total, 73 caractéristiques de la VFC et 18 caractéristiques de la VFR dans les domaines temporel et non linéaire sont extraites. Les caractéristiques fréquentielles ne sont pas extraites en raison de la courte durée de la fenêtre.

### 2.3.2 Sélection des caractéristiques

Le critère de Fisher est utilisé comme méthode de sélection des caractéristiques. Il consiste à calculer un score  $F$  pour chaque caractéristique, en comparant la variance entre les classes à la variance à l'intérieur des classes. Les caractéristiques ayant le score  $F$  le plus élevé sont considérées comme les plus importantes. Le score  $F$  est défini par :

$$F = \frac{\sum_{k=1}^K n_k (\mu_k - \mu)^2}{\sum_{k=1}^K n_k \sigma_k^2} \quad (1)$$

où  $K$  est le nombre de classes,  $n_k$ ,  $\mu_k$  et  $\sigma_k$  sont le nombre d'observations, la moyenne et l'écart-type de la caractéristique au sein de la classe  $k$ , et  $\mu$  est la moyenne globale de la caractéristique. Le paramètre  $K$  est fixé à 2 pour classer les segments ECG et respiratoires en segments réguliers et non réguliers.

Pour chaque enregistrement, les caractéristiques VFC et VFR sont classées séparément par ordre décroissant d'importance. Les caractéristiques finalement retenues sont celles qui sont classées dans le "top 5" pour au moins 50% des enregistrements. Ceci conduit à la sélection de six caractéristiques, qui sont présentées dans la Table 1.

### 2.3.3 Classification non supervisée K-means

Après l'extraction et la sélection des caractéristiques, le partitionnement automatique des segments est effectué par l'algorithme de clustering *K-means* de la librairie Python *scikit-learn*. C'est un algorithme non supervisé qui partitionne les données en  $K$  clusters, basé sur trois étapes. La première est l'initialisation des centroïdes des clusters à l'aide d'une méthode intégrée dans la librairie permettant d'optimiser la convergence (*k-means++*). Dans la deuxième étape, chaque segment est assigné à l'un des clusters, sur la base de la distance la plus courte calculée. La troisième étape consiste à mettre à jour les centroïdes en calculant la valeur moyenne de tous les segments assignés à chaque centroïde précédent. L'algorithme itère les deux dernières étapes jusqu'à ce que la différence calculée entre l'ancien et le nouveau centroïde soit inférieure à un seuil prédéfini.

Afin de garantir l'attribution du label "1" au cluster SC, on exploite la connaissance a priori qu'en SC l'intervalle RR est plus long et que la coordonnée *CVBB* a une valeur plus faible. Compte tenu de cet élément, l'ordre des centroïdes des clusters est soit conservé, soit inversé.

Enfin, deux approches de clustering sont comparées : i) l'une utilisant un seul *K-means* pour l'ECG et la respiration (C1), et ii) l'autre utilisant deux *K-means*, un pour l'ECG et un pour la respiration (C2).

TABLE 1 : Caractéristiques retenues pour la classification.

Type	Caractéristique	Description	Occurrence
VFR temporel	$T_{Insp}$	La durée moyenne de l'inspiration	50%
	$CVBB$	L'écart-type des intervalles BB (SDBB) divisé par la moyenne des intervalles BB (MeanBB)	92%
	$MCVBB$	L'écart absolu médian des intervalles BB (MadBB) divisé par la médiane des intervalles BB (MedianNN)	84%
VFC temporel	$Prc20RR$	Le 20e percentile des intervalles RR	64%
VFC non linéaire	$FuzzyEN$	La fuzzy entropy	68%
	$KFD$	La dimension fractale de Katz	60%

## 2.4 Fusion des trois modalités

Les résultats de la segmentation du mouvement et les signaux binaires résultant du clustering des signaux physiologiques sont ensuite fusionnés pour estimer le SC. Deux stratégies sont proposées : i) utiliser la segmentation du mouvement (0/1), classifier tous les segments physiologiques, et réaliser la fusion avec un "ET" (F1), ou ii) affecter d'abord tous les intervalles de mouvement à l'état NSC, puis finaliser la classification NS/NSC avec les segments physiologiques observés lors des intervalles de non-mouvement (F2).

## 3 Résultats

Cette section est divisée en trois parties. Tout d'abord, une étude préliminaire vise à évaluer la contribution de chaque modalité dans l'estimation du SC, et à identifier la combinaison optimale. Ensuite, les performances sont présentées en fonction des différentes stratégies de clustering des signaux physiologiques et de fusion avec le mouvement. Enfin, les résultats sont analysés en fonction de l'âge des bébés.

Sur l'ensemble des enregistrements, le temps passé en SC est plus faible ( $23.9 \pm 7.76\%$ ) qu'en NSC ( $76.1 \pm 7.76\%$ ). Pour tenir compte de ce déséquilibre, les performances sont évaluées par la sensibilité, la spécificité, le balanced accuracy et le Cohen's kappa (classiquement utilisé pour évaluer l'accord entre deux annotations [10]). Enfin, dans la perspective clinique d'un système auto-apprenant et spécifique bébé, les performances sont calculées pour chaque bébé puis moyennées.

### 3.1 Etude préliminaire

Afin d'identifier les contributions des différentes modalités (mouvement, ECG et respiration), celles-ci ont d'abord été traitées séparément. Cela a conduit à trois estimations différentes du SC ( $\widetilde{SC}$ ) : i) en sortie du mouvement, ii) en sortie du  $K$ -means appliqué à l'ECG et iii) en sortie du  $K$ -means appliqué à la respiration. Puis, les différents  $\widetilde{SC}$  ont été fusionnés avec un "ET", pour étudier les combinaisons de deux ou trois modalités. La Figure 2 illustre les résultats obtenus pour les sept configurations possibles. On observe d'abord que le signal de non-mouvement est très sensible ( $0.98 \pm 0.03$ ), détectant presque tous les états SC, mais a une faible spécificité ( $0.25 \pm 0.12$ ). Les signaux ECG et respiratoires ont une sensibilité plus faible

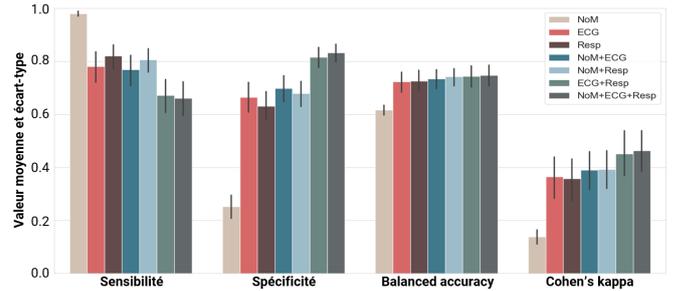


FIGURE 2 : Comparaison des performances des différentes combinaisons de modalités pour l'estimation du SC.

( $0.78 \pm 0.14$  et  $0.82 \pm 0.12$ ), mais ont une spécificité plus élevée ( $0.66 \pm 0.14$  et  $0.63 \pm 0.14$ ), respectivement. Ainsi, en combinant les signaux, la sensibilité diminue mais la spécificité, le balanced accuracy et le Cohen's kappa augmentent. Les performances ainsi obtenues confirment que le meilleur compromis est obtenu par la combinaison des trois modalités. Cette dernière conduit à une sensibilité de  $0.66 \pm 0.17$ , une spécificité de  $0.83 \pm 0.09$ , un balanced accuracy de  $0.75 \pm 0.10$  et un Cohen's kappa de  $0.46 \pm 0.2$ .

### 3.2 Estimation du sommeil calme

Pour l'estimation du SC, deux stratégies de clustering des données physiologiques (section 2.3.3) et deux stratégies de fusion avec le mouvement ont été étudiées (section 2.4). Ceci conduit à quatre approches différentes, dont les performances sont données dans la Table 2. La méthode C2-F1 correspond au cas "NoM+ECG+Resp" déjà présenté dans la section 3.1.

TABLE 2 : Comparaison des méthodes de clustering et fusion.

Clustering	Fusion	Balanced accuracy	Cohen's kappa
C1	F1	$0.76 \pm 0.09$	$0.45 \pm 0.22$
C2	F1	$0.75 \pm 0.10$	$0.46 \pm 0.20$
C1	F2	$0.76 \pm 0.10$	$0.50 \pm 0.23$
C2	F2	$0.72 \pm 0.10$	$0.46 \pm 0.21$

Les résultats obtenus montrent que les performances sont légèrement meilleures avec la méthode C1-F2, c'est-à-dire lorsque l'on affecte d'abord les intervalles de mouvement à l'état NSC, puis qu'on applique le clustering aux segments ECG et de respiration correspondant aux intervalles de non-mouvement, et en utilisant un seul classifieur.

### 3.3 Évaluation stratifiée en fonction de l'APM

L'expérimentation a montré que les performances varient avec l'APM des bébés. La Figure 3 montre les résultats obtenus (avec la méthode C1-F2) lorsque les enregistrements des différents groupes et jours sont analysés séparément.

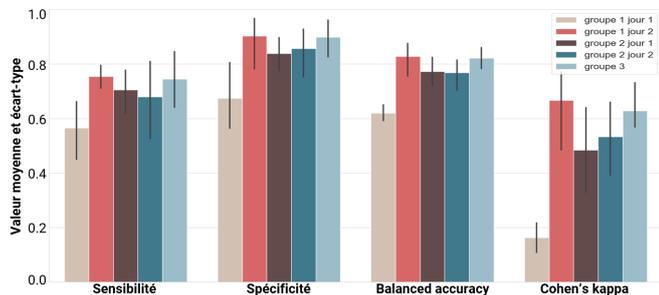


FIGURE 3 : Comparaison des performances d'estimation du sommeil calme en fonction l'APM.

On observe une homogénéité des résultats dans les quatre derniers sous-groupes, mais pour les bébés du groupe 1 enregistré le jour 1, les performances obtenues sont plus faibles. Autrement dit, pour les bébés avec un APM supérieur à 33 semaines, la méthode fournit des résultats intéressants avec un balanced accuracy moyen de 0.80 et un Cohen's kappa moyen de 0.58. Pour les grands prématurés enregistrés juste après la naissance (APM < 30 semaines), les performances moins bonnes sont sans doute liées au fait que, chez ces bébés, les périodes de SC sont rares et plus courtes [4], et donc sans doute plus difficiles à détecter. De plus, chez ces bébés, l'information contenue dans l'ECG et dans la respiration est a priori moins riche, du fait de leur grande immaturité.

## 4 Discussion et Conclusion

Dans cet article, une méthode a été proposée pour estimer le sommeil calme chez les prématurés et les nouveau-nés à terme. La nouveauté de cette approche réside dans l'utilisation d'une approche non supervisée pour identifier automatiquement les périodes de l'ECG et du signal respiratoire qui correspondent au SC et au NSC. Les indicateurs binaires issus de cette étape de clustering ont été complétés par des informations sur le mouvement, extraites de la vidéo. Toutes les combinaisons de ces trois modalités ont été étudiées et les meilleurs résultats ont été obtenus lorsque les trois signaux étaient pris en compte. Ensuite différentes stratégies de clustering et de fusion ont été comparées. Les meilleurs performances ont été obtenues lorsque l'on affecte d'abord les intervalles de mouvement à l'état NSC, puis qu'on utilise un seul classifieur global pour les segments ECG et de respiration. Enfin, nous avons montré que les performances dépendent de l'APM, avec des résultats inférieurs obtenus avec les grands prématurés, enregistrés juste après la naissance.

Ces résultats sont encourageants et permettent d'obtenir une première estimation du sommeil calme par l'utilisation des signaux de mouvement et cardio-respiratoires, avec une approche non supervisée. Cette approche peut être utilisée en continu dans l'USIN, en particulier pour le suivi des prématurés, sans nécessiter un système de polysomnographie. Elle ouvre la perspective de suivi du SC du prématuré sans intervention manuelle. C'est la première fois, à notre connaissance,

qu'une telle approche est proposée. En comparaison avec les approches de la littérature qui sont basées sur l'apprentissage classique et l'apprentissage profond, notre méthode montre des performances meilleures (par exemple, le Cohen's kappa est égal à 0.27 en moyenne dans [10]). Plus important encore, ce travail a pour caractéristique d'avoir été réalisé sur des données acquises en conditions cliniques réelles.

Les développements futurs concernent i) les aspects méthodologiques tels que l'introduction d'une loi de fusion optimale en intégrant un processus de commutation en cas de données manquantes dans une modalité, ii) les aspects cliniques avec l'estimation de tous les stades de sommeil des nouveau-nés et iii) une étude clinique de grande envergure montrant le bénéfice pour le bébé de suivre le SC en USIN.

## Références

- [1] TB BRAZELTON et JK NUGENT : Neonatal behavioral assessment scale. Cambridge University Press, 1995.
- [2] S. CABON, F. PORÉE, A. SIMON et al. : Audio-and video-based estimation of the sleep stages of newborns in neonatal intensive care unit. Biomed Signal Process Control, 52:362–70, 2019.
- [3] S. CABON, R. WEBER, S. SIMON et al. : Functional age estimation through neonatal motion characterization using continuous video recordings. IEEE J Biomed Health Inform, 27:1500–11, 2023.
- [4] L. CAILLEAU, R. WEBER, S. CABON et al. : Quiet sleep organization of very preterm infants is correlated with postnatal maturation. Front Pediatr, 8, 2020.
- [5] M. DOYEN, D. GE, A. BEUCHÉE et al. : Robust, real-time generic detector based on a multi-feature probabilistic method. Plos One, 14(10):e0223785, 2019.
- [6] J. ISLER, T. THAI, M. MYERS et W. FIFER : An automated method for coding sleep states in human infants based on respiratory rate variability. Dev Psychobiol, 58(8):1108–15, 2016.
- [7] S. KHREIS, D. GE, H.A. RAHMAN et G. CARRAULT : Breathing rate estimation using Kalman smoother with electrocardiogram and photoplethysmogram. IEEE Trans Biomed, 67(3):893–904, 2019.
- [8] D. MAKOWSKI, T. PHAM, Z. LAU et al. : Neurokit2 : A python toolbox for neurophysiological signal processing. Behav Res Methods, 53(4):1689–96, 2021.
- [9] Digi-NewB GCS HUGO CHU monitoring SYSTEM : <http://www.digi-newb.eu>. 14 April 2020.
- [10] T. SENTNER, X. WANG, E. de GROOT et al. : The sleep well baby project : an automated real-time sleep-wake state prediction algorithm in preterm infants. Sleep, 45(10):zsac143, 2022.
- [11] J. WERTH, M. RADHA, P. ANDRIESEN et al. : Deep learning approach for ECG-based automatic sleep state classification in preterm infants. Biomed Signal Process Control, 56:101663, 2020.