

Approche bayésienne pour l'extraction et la classification de potentiels d'action

Steven LE CAM¹, Radu RANTA¹

¹Centre de Recherche en Automatique de Nancy (CRAN), CNRS UMR 7039 - Université de Lorraine
1, av. de la forêt de Haye, 54500 Vandoeuvre-les-Nancy, France
steven.le-cam@univ.lorraine.fr

Résumé – Une approche bayésienne pour l'extraction et la classification des potentiels d'action dans les signaux EEG de microélectrodes est proposée. L'originalité de la méthode réside dans la séparation des formes de potentiel d'action du signal de fond (communément appelé LFP) dans le but d'améliorer le résultat de classification. Usuellement cette séparation est obtenue par un simple filtrage passe-haut, les potentiels d'action étant supposés présenter une activité essentiellement hautes fréquences alors que le LFP se situe dans les fréquences basses. Le LFP est ici modélisé par un processus gaussien coloré, et un mélange de gaussiennes multivariées est utilisé pour modéliser la distribution des formes d'onde des potentiels d'action. L'optimisation du modèle conduit à une estimation *a posteriori* de chaque forme d'onde individuelle des potentiels d'action détectés, une version du signal LFP séparée de ces potentiels d'action, ainsi qu'à une classification des potentiels d'action ainsi estimés. Cette classification est menée dans le domaine temps/fréquence par décomposition en ondelettes, et les performances sont évaluées sur une base de données simulées. La méthode est comparée à des méthodes récentes de la littérature, ainsi qu'à une version précédente de la méthode où la classification était opérée sur les échantillons temporels des formes d'onde extraites du signal.

Abstract – A Bayesian approach for the extraction and classification of action potentials in microelectrode EEG signals is proposed. The originality of the method lies in the separation of the action potential forms from the background signal (commonly called LFP) in order to improve the classification result. Usually this separation is obtained by a simple high-pass filtering, the action potentials being supposed to hold mainly high frequency components while the LFP is located in the low frequency band. The LFP is modeled by a colored Gaussian process, and a mixture of multivariate Gaussians is used to model the waveform distribution of the action potentials. The optimization of the model leads to the *posterior* estimate of each individual waveform of the detected action potentials, a version of the LFP signal with these action potentials removed, as well as a classification of the estimated action potential waveforms. This classification is carried out in the wavelet domain, and its performance is evaluated on a simulated data base. The method is compared to recent algorithms of the literature, as well as to a former version of the method where the clustering was carried out directly from the time samples of the estimated action potential waveforms.

1 Introduction

L'enregistrement de signaux électrophysiologiques à l'échelle du neurone chez les humains est une pratique de plus en plus répandue, rendue possible par l'implantation d'électrodes sous durales (grille d'électrodes sur le cortex) ou intra-cérébrales (implantation d'électrode en aiguille dans le volume cérébral). Ces données donnent accès à de précieuses informations sur les mécanismes cérébraux cognitifs ou pathologiques à l'échelle de la cellule neuronale. La projection des potentiels d'action (ou spikes) générées par un neurone sur une électrode dépend de la morphologie de ce neurone et de sa position relative par rapport à l'électrode [1]. Il est donc possible d'identifier une signature associée à chaque neurone et ainsi de séparer les trains de spikes des différents neurones dont les activités sont visibles sur une électrode donnée. Cette étape de détection et de classification est connue sous le nom de spike sorting, et est à la base de l'analyse de ces signaux permettant d'aborder des questions neuroscientifiques fondamentales [2, 3].

L'activité visible sur les électrodes de type micro est largement considérée comme étant une somme de contributions

macroscopiques de population de neurones synchronisés situés jusqu'à plusieurs millimètres de distance (le LFP), sur laquelle s'additionne la projection des spikes issus de neurones positionnés jusqu'à $200\mu m$ de l'électrode. Parmi les spikes qui peuvent être détectés dans le signal brut, on distingue alors les spikes pour lesquels le rapport signal à bruit est suffisant sur la mesure (i.e. qui ressortent suffisamment du signal LFP), et dont la séparation en classe des formes d'onde est possible, de ceux noyés dans le LFP et de ce fait pour lesquels une classification n'est pas possible : on parle alors de single unit (SU) dans le premier cas, et de classe multi units (MU) dans le second cas. La séparation des formes d'onde des spikes et du LFP est donc une étape importante vers l'obtention de résultats de classification fiable, cette étape étant généralement menée par filtrage [4, 5, 6]. La partie basse-fréquence (généralement en deçà de 300Hz) est associée au LFP, alors que l'activité des spikes se situe principalement en haute fréquence. L'objectif de ce travail est d'améliorer la séparation spikes/LFP afin de restituer les composantes basses fréquences des spikes, dans le but d'améliorer leur séparabilité et ainsi le résultat de classification.

La reconstruction des formes d'onde, leur séparation du LFP et leur classification est un problème inverse dont nous proposons une résolution par approche bayésienne. Si certains systèmes d'acquisition permettent un traitement multi-canal pour la classification du fait de la proximité des capteurs entre eux (grille d'électrodes), nous nous plaçons ici dans le cas de l'utilisation d'électrodes Benhke-Fried [7] où la distance entre capteurs n'est pas connue et peut aller jusqu'à plusieurs millimètres, ce qui impose un traitement mono-canal de l'information. Dans ce cadre, la composante associée au LFP est modélisée par un processus gaussien coloré de moyenne nulle, et les formes d'ondes sont supposées suivre un mélange de gaussiennes multivariées. Dans une première version de la méthode [8], les échantillons temporels des spikes étaient considérés. Nous proposons ici de plonger l'étape de classification dans l'espace temps/fréquence par décomposition en ondelettes, et nous montrons le gain apporté par la sélection de coefficients discriminants pour la classification. Les distributions *a posteriori* du LFP ainsi que de chaque forme d'ondes sont estimées à l'aide d'un algorithme d'optimisation variationnelle, et l'identification des labels de chaque classe est donné par une décision du maximum *a posteriori* (MAP). La méthode est testée sur une base de données simulées [9] déjà utilisée précédemment pour la validation et la comparaison de méthodes de spike sorting [6, 10].

2 Modèle génératif

Soit y le signal enregistré par l'électrode de longueur T échantillons. Nous supposons le modèle additif suivant :

$$y = x + S + \epsilon \quad (1)$$

où x est le LFP, S l'activité des N formes d'onde de spikes enregistrées sur l'électrode, et ϵ un bruit additif.

S peut-être écrit comme la somme des N formes d'ondes de spikes s_n détectés au préalable dans le signal (voir section 4), chaque s_n étant localisée sur un support temporel restreint T_s :

$$S = \sum_{n=1}^N s_n * \delta_n \quad (2)$$

δ_n étant un dirac permettant de positionner s_n sur son support temporel correspondant dans y . Les N formes d'onde s_n détectés dans le signal sont supposées être issues d'une distribution gaussienne multivariée. Une incertitude est ainsi introduite sur la forme des ondes issues de chaque classe (ou neurone), liée aux variations inhérentes aux processus biologiques mesurés (effet d'épuisement, mouvement des tissus...) et au bruit d'acquisition. Afin de modéliser les classes de spikes, le processus caché Z est introduit dans le modèle, prenant ses valeurs dans l'ensemble $\{\lambda_1, \dots, \lambda_K\}$, λ_k représentant la classe des formes d'onde des spikes associées au k^{th} neurone. Le processus s_n suit alors un modèle de mélange de gaussiennes multivariée (GMM) :

$$p(s_n) = \sum_{k=1}^K \pi_k p(s_n | z_n = \lambda_k) \quad (3)$$

$$p(s_n | z_n = \lambda_k) = \mathcal{N}(s_n | s_k, \Sigma_k) \quad (4)$$

avec $\pi_k = p(z = \lambda_k)$ les probabilités *a priori* sur Z , et $\{s_k, \Sigma_k\}$ les paramètres de moyenne et de covariance pour la classe λ_k .

Comme proposé par Zanos *et al.* [11], x suit une distribution gaussienne $\mathcal{N}(0, \gamma \Sigma_x)$, où γ est un paramètre d'échelle associé à la puissance du signal LFP¹. Finalement, ϵ est un processus gaussien aléatoire centré de variance σ_ϵ^2 . La vraisemblance des données s'écrit :

$$p(y|x, S) \propto \exp -\frac{1}{2\sigma_\epsilon^2} \|y - x - S\|_2^2 \quad (5)$$

On en déduit la log-vraisemblance \mathcal{L} du modèle :

$$\begin{aligned} \mathcal{L} &= \log p(y, x, S, Z) \\ &= \log p(y|x, S) + \log p(x) + \log p(S|Z) \\ &\quad + \log p(Z) \end{aligned} \quad (6)$$

L'optimisation de l'ensemble des paramètres du modèle $\Theta = \{\gamma, \{s_k, \Sigma_k\}_{k=1}^K, \sigma_\epsilon^2\}$ est menée à l'aide d'une procédure bayésienne variationnelle décrite dans la prochaine section.

3 Optimisation variationnelle

L'optimisation variationnelle consiste en l'introduction des lois candidates $q(x, S, Z) = q(x) \prod_{n=1}^N q(s_n) q(z_n)$, approximation de la vraie densité *a posteriori* $p(x, S, Z|y)$. Deux étapes d'estimation et de maximisation sont itérés jusqu'à convergence comme suit :

Etape d'estimation La première étape consiste en la mise à jour des $2N + 1$ $q(\cdot)$ facteurs constitutifs de cette loi candidate : chaque facteur $q(\cdot)$ est donné par l'espérance de la log-vraisemblance du modèle \mathcal{L} conditionnellement aux $2N$ $q(\cdot)$ facteurs restants [12]. En particulier pour le modèle choisi ci-dessus, les densités $q(x)$ and $q(s_n)$ sont des gaussiennes multivariées, respectivement de moyennes \hat{x} et \hat{s}_n et de matrice de covariance $\hat{\Sigma}_x$ et $\hat{\Sigma}_n$:

$$\hat{x} = \gamma \Sigma_x (\gamma \Sigma_x + \sigma_\epsilon^2 I_T)^{-1} (y - \hat{S}) \quad (7)$$

$$\hat{\Sigma}_x = \sigma_\epsilon^2 \gamma \Sigma_x (\gamma \Sigma_x + \sigma_\epsilon^2 I_T)^{-1} \quad (8)$$

$$\hat{s}_n = \hat{\Sigma}_n \left(\frac{1}{\sigma_\epsilon^2} \mathcal{S}_n (y - \hat{x}) + \sum_{k=1}^K \hat{\pi}_{nk} \Sigma_k^{-1} s_k \right) \quad (9)$$

$$\hat{\Sigma}_n = \left(\frac{1}{\sigma_\epsilon^2} I_{T_s} + \sum_{k=1}^K \hat{\pi}_{nk} \Sigma_k^{-1} \right)^{-1} \quad (10)$$

où \hat{S} contient les N moyennes *a posteriori* des formes d'onde \hat{s}_n à leur position respective dans le signal, et $\mathcal{S}_n(\cdot)$ est la restriction au support de s_n de longueur T_s dans y . $\hat{\pi}_{nk} = q(z_n =$

1. seul γ sera ré-estimé par la méthode, Σ_x est fixée à l'initialisation (voir section 4)

λ_k) sont les estimés des probabilités *a posteriori* pour chaque label z_n . La classification des formes d'onde à partir des échantillons temporels de \hat{s}_n implique une discrimination de ces formes dans un espace de dimension égal à T_s , soit plusieurs dizaines d'échantillons. Afin d'augmenter le pouvoir discriminant de la méthode, on propose ici de décomposer les spikes en ondelettes et d'en sélectionner le coefficient les plus pertinents (les M coefficients de plus grande variance dans la décomposition des N formes d'ondes détectées sont retenus). On note \hat{s}_n^w les coefficients sélectionnés pour l'onde \hat{s}_n , et on définit $\|s\|_{\Sigma}^2 = s^T \Sigma^{-1} s$. Les facteurs $\hat{\pi}_{nk}$ sont donnés par :

$$\hat{\pi}_{nk} \propto \frac{\pi_k}{|\Sigma_k^w|} \exp\left(-\frac{1}{2}\|\hat{s}_n^w - s_k^w\|_{\Sigma_k^w}^2 - \frac{1}{2}\text{Tr}[\hat{\Sigma}_n^w \Sigma_k^{w-1}]\right) \quad (11)$$

au gain de normalisation à 1 près².

Etape de maximisation Les paramètres du modèle θ sont obtenus par maximisation de l'espérance *a posteriori* de la log-vraisemblance du modèle :

$$\gamma = \frac{1}{T} \left(\|\hat{x}\|_{\Sigma_x}^2 + \text{Tr}[\hat{\Sigma}_x \Sigma_x^{-1}] \right) \quad (12)$$

$$\sigma_\epsilon^2 = \frac{1}{T} \left(\|y - \hat{x} - \hat{S}\|_2^2 + \text{Tr}[\hat{\Sigma}_x] + \sum_{n=1}^N \text{Tr}[\hat{\Sigma}_n] \right) \quad (13)$$

$$s_k = \frac{\sum_{n=1}^N \hat{\pi}_{nk} \hat{s}_n}{\sum_{n=1}^N \hat{\pi}_{nk}} \quad (14)$$

$$\Sigma_k = \frac{\sum_{n=1}^N \hat{\pi}_{nk} \left(\|\hat{s}_n - s_k\|_2^2 + \hat{\Sigma}_n \right)}{\sum_{n=1}^N \hat{\pi}_{nk}} \quad (15)$$

4 Initialisation de la méthode

L'approche proposée nécessite au préalable la détection des spikes à classer, l'initialisation de la matrice de covariance du LFP, ainsi que du nombre de classe à estimer.

La **détection des spikes** est effectuée par seuillage du signal brut filtré dans une bande pré-déterminée, méthodologie largement usitée dans la littérature [6, 4, 3]. Le seuil est donné par un multiple d'une estimation robuste de l'écart-type du bruit basé sur la médiane du signal brut [13, 4]. Au sein d'une fenêtre de $3ms$, seul le maximum en valeur absolu est retenu comme le pic d'un évènement de spike, le recouvrement entre évènements n'étant pas considéré par la méthode.

Matrice de covariance du LFP. Nous avons posé un *a priori* de type gaussien multivarié de moyenne nulle sur le LFP x . La matrice de covariance $\gamma \Sigma_x$ de ce processus est de très grande taille ($T \times T$), son estimation ainsi que le calcul explicite des mises à jour de la moyenne et de la covariance *a posteriori* du processus x (eq. (7),(8)) ne sont pas faisable en pratique. Nous supposons la matrice Σ_x circulante [11], et nous profitons de

cette propriété pour transposer les calculs en fréquence et proposer une estimation basée sur la densité spectrale de puissance (DSP) du signal³. Après passage dans le domaine de Fourier, les équations (7) et (8) peuvent se calculer par des opérations de convolution pour lesquels des algorithmes rapides sont disponibles [11, 8]. La DSP du LFP x est estimé à partir de celle du signal brut y . Zanos *et al.* [11] pose un modèle paramétré de type exponentiel $-exp(1 + \log(f))$, ajusté sur une représentation dans l'espace log-log de la DSP du signal brut dans l'intervalle 1Hz à 150Hz, i.e. la partie du spectre associée au LFP. La DSP est ensuite extrapolée à partir de ce modèle dans les hautes fréquences. C'est donc sur ce modèle que repose l'estimation de Σ_x dans l'espace de Fourier, ce paramètre est fixé à l'initialisation et seul le facteur d'échelle γ est ré-estimé afin d'ajuster la puissance du LFP et du bruit σ_ϵ^2 à partir des données.

La **détermination du nombre de classes K** à estimer est menée en amont de la méthode à l'aide d'un algorithme GMM-EM appliqué pour différentes valeurs de K sur les M coefficients issus de la décomposition des spikes extraits du signal brut après filtrage passe-bande. La valeur de K pour laquelle un critère de type MDL est maximisé est finalement retenue. Ce premier résultat de classification permet également l'initialisation des **moeyennes et covariances des classes**.

5 Résultats

Nous évaluons la performance de notre algorithme sur le jeu de données de Camunas *et al.* [9], qui fournit une base de signaux réalistes (<https://www.135.lamp.le.ac.uk/hgr3/>) : elle contient 5 signaux d'une durée de 10 minutes pour des nombres de SU simulés allant de 2 à 20, soit $5 \times 19 = 95$ signaux au total. Sur chaque signal est ajouté la projection de spikes considérés comme non classables (MU). Cette base de données a été utilisée dans plusieurs études récentes pour évaluer les approches de spike sorting [6, 10]. La méthode est évaluée en utilisant les mêmes critères que ceux utilisés dans ces références : un cluster est considéré comme un vrai positif s'il contient plus de 50% des spikes d'un SU simulé, et si au moins 50% des spikes provenant de cette unité sont classés dans ce cluster. Sur la base de ce critère, si un cluster ne correspond à aucun SU et n'est pas formé d'au moins 50% d'évènements simulés de type MU, il est considéré comme un faux positif.

Nous comparons notre algorithme (dans ses versions en ondelettes et temporelle [8]) avec trois algorithmes récents de la littérature : Klusta [5], Combinato [6], Waveclus dans sa version automatique [10], exécutés depuis l'outil Spikeinterface [14], une plateforme en libre accès mettant plusieurs spike sorters à disposition. Ces trois méthodes considèrent toutes la classification des formes d'onde extraites du signal après filtrage passe-bande. Un filtre du quatrième ordre dans la bande [300-3000]Hz a été considéré [10], et la détection des spikes est effectuée en utilisant le seuillage robuste décrit plus haut.

2. Les paramètres de moyenne s_k^w et de covariances Σ_k^w et $\hat{\Sigma}_n^w$ dans l'espace ondelettes sont obtenus par transposition de s_k^w , Σ_k et $\hat{\Sigma}_n$ sur la base de décomposition choisie, puis restriction aux termes associés aux coefficients sélectionnés

3. Cette propriété n'est plus vraie après passage en ondelettes, ce qui ne nous permet pas en pratique de déployer l'ensemble de la méthode dans cet espace temps/fréquence

Le facteur multiplicatif pour le seuillage robuste a été fixé à 4.5, meilleur compromis entre la détection des événements de spikes simulés (SU ou MU) et les fausses détections. En moyenne, 75.7% + -7.4% du nombre total de spikes sont détectés (MU inclus), 96,5% + -1.9% des SU sont détectés, tandis que le nombre de faux positifs (événement détecté ni SU ni MU) varie de 1,2% à 3,4% du nombre total de spikes détectés (soit de 5 à 123 faux positifs selon le signal considéré). Les résultats sont synthétisés figure 1, en terme de vrai et faux positifs, et montrent clairement que la méthode proposée est supérieure aux autres algorithmes testés pour retrouver les clusters, en particulier lorsque le nombre de SU à estimer est élevé. En ce qui concerne les faux-positifs, l’algorithme proposé est également très performant et produit beaucoup moins de faux clusters que ses concurrents.

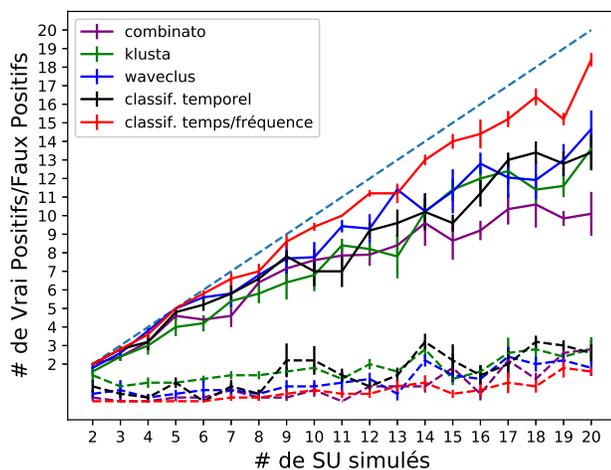


FIGURE 1 – Performance comparative de différents algorithmes (Combinato, Klusta, Waveclus, et l’algorithme proposé dans sa version temporelle [8] et temps/fréquence) sur la base de donnée de Camunas *et al.* [9]. En trait plein : nombre de bonne détection en fonction du nombre de neurones simulés (moyenne et SEM sur 5 simulations pour chaque nombre de SU simulé). En trait discontinu : faux-positifs.

6 Conclusion

Nous proposons une approche bayésienne pour la séparation spikes/LFP et la classification simultanés des spikes, cette dernière étant menée dans l’espace temps/fréquence par décomposition en ondelettes. Les performances de cette méthodologie sont illustrées sur une base de signaux simulés. Une étape de validation sur signaux réels est requise afin d’éprouver la méthode sur des données non synthétiques. Dans ce cas nécessairement plus complexe, des étapes de pré-traitement pour rejeter les fausses détections sont nécessaires, et des critères supplémentaires basés sur des *a priori* bio-physiques [15] peuvent être utilisés pour écarter les résultats de classification peu vraisemblables. Enfin, au delà de la classification des spikes, le second objectif de la méthode est de fournir une estimation du LFP dénués de l’activité des spikes détectés, cet aspect reste à être validé et comparé aux méthodes existantes [11].

Références

- [1] H. Tran, R. Ranta, *et al.*, “Fast simulation of extracellular action potential signatures based on a morphological filtering approximation,” *Journal of computational neuroscience*, vol. 48, no. 1, pp. 27–46, 2020.
- [2] G. Buzsáki, “Large-scale recording of neuronal ensembles,” *Nature neuroscience*, vol. 7, no. 5, p. 446, 2004.
- [3] H. G. Rey, C. Pedreira, and R. Q. Quiroga, “Past, present and future of spike sorting techniques,” *Brain research bulletin*, vol. 119, pp. 106–117, 2015.
- [4] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, “Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering,” *Neural computation*, vol. 16, no. 8, pp. 1661–1687, 2004.
- [5] C. Rossant, S. N. Kadir, *et al.*, “Spike sorting for large, dense electrode arrays,” *Nature neuroscience*, vol. 19, no. 4, pp. 634–641, 2016.
- [6] J. Niediek, J. Boström, *et al.*, “Reliable analysis of single-unit recordings from the human brain under noisy conditions : tracking neurons over hours,” *PloS one*, vol. 11, no. 12, p. e0166598, 2016.
- [7] I. Fried, C. L. Wilson, *et al.*, “Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients,” *Journal of neurosurgery*, vol. 91, no. 4, pp. 697–705, 1999.
- [8] S. Le Cam, H. Tran, *et al.*, “A bayesian approach for simultaneous spike extraction and sorting,” in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2019, pp. 993–996.
- [9] L. A. Camuñas-Mesa and R. Q. Quiroga, “A detailed and fast model of extracellular recordings,” *Neural computation*, vol. 25, no. 5, pp. 1191–1212, 2013.
- [10] F. J. Chaure, H. G. Rey, and R. Quiroga, “A novel and fully automatic spike-sorting implementation with variable number of features,” *Journal of neurophysiology*, vol. 120, no. 4, pp. 1859–1871, 2018.
- [11] T. P. Zanos, P. J. Mineault, and C. C. Pack, “Removal of spurious correlations between spikes and local field potentials,” *Journal of neurophysiology*, vol. 105, no. 1, pp. 474–486, 2010.
- [12] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
- [13] D. L. Donoho and J. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [14] A. P. Buccino, C. L. Hurwitz, *et al.*, “Spikeinterface, a unified framework for spike sorting,” *Elife*, vol. 9, p. e61834, 2020.
- [15] D. N. Hill, S. B. Mehta, and D. Kleinfeld, “Quality metrics to accompany spike sorting of extracellular signals,” *Journal of Neuroscience*, vol. 31, no. 24, pp. 8699–8705, 2011.