

Généralisation des micro-états EEG par apprentissage régularisé temporellement de dictionnaires topographiques

Yoann ISAAC^{1,2}, Quentin BARTHÉLEMY³, Cédric GOUY-PAILLER¹, Jamal ATIF⁴, Michèle SEBAG²

¹CEA, LIST, 91191 Gif-sur-Yvette, France

²TAO, CNRS – INRIA – LRI, 91405 Orsay, FRANCE

³Mensia Technologies, ICM, Hôpital de la Pitié-Salpêtrière, 75013 Paris, FRANCE

⁴PSL, Université Paris-Dauphine, LAMSADE, CNRS, UMR 7243, 75775 PARIS, FRANCE

prénom.nom@{cea.fr, lri.fr, mensiatech.com, dauphine.fr}

Résumé – Le modèle des micro-états décrit les signaux EEG par des suites de topographies associées à des états cérébraux demeurant stables durant quelques dizaines de millisecondes. La généralisation proposée dans cet article considère un modèle redondant autorisant plusieurs états à être actifs simultanément. Un apprentissage de dictionnaire régularisé temporellement est proposé afin d’extraire ces états. L’efficacité de représentation des deux modèles est comparée sur des signaux de synthèse ainsi que sur des signaux réels pour l’étude du potentiel évoqué P300.

Abstract – The microstate model describes EEG signals as series of topographies remaining stable during several tens of milliseconds and associated to brain states. The proposed generalization in this paper is based on an overcomplete model allowing several states to be active simultaneously. A dictionary learning algorithm with a temporal regularization is proposed to extract these states. The representation effectiveness of both models is compared on artificial and real signals for the extraction of the evoked potential P300.

1 Introduction

L’électroencéphalographie (EEG) permet, au moyen d’un ensemble d’électrodes disposées à la surface du crâne, de mesurer une partie de l’activité électrique du cerveau. Les signaux résultants peuvent être vus comme une succession de topographies, dont chacune d’elles est constituée par l’agrégation spatiale, à un instant donné, des potentiels enregistrés par les électrodes. Ces topographies présentent des dynamiques complexes qui dépendent en partie des tâches cérébrales réalisées. Bien que l’interprétation complète de leur évolution soit difficile, certaines études [1] ont mis en évidence une alternance entre des périodes de stabilité des topographies, d’une durée de quelques dizaines de millisecondes, et des périodes de transition relativement brusques. Les topographies observées pendant les phases de stabilité sont appelées micro-états et leur enchaînement peut être interprété comme la réalisation d’une succession de tâches cérébrales [2].

Des études statistiques réalisées sur des suites temporelles de tels états ont lié les modifications de leurs ordres d’apparition et/ou de leurs durées à des maladies mentales comme la schizophrénie [3] ou la dépression [4]. Ils ont également permis la caractérisation des signaux EEG associés aux états de repos d’un sujet [5] et ont été considérés pour l’étude de potentiels évoqués [6, 7], où ce modèle présente des avantages par rapport aux approches d’analyse classiques [8].

Dans le modèle des micro-états, chaque topographie d’un si-

gnal EEG est associée à un état et l’évolution temporelle de ces états est une fonction constante par morceaux. L’extraction des micro-états est donc un problème de quantification vectorielle (QV) contrainte temporellement : *i.e.*, un problème où l’on cherche à coder un ensemble de mesures successives (topographies) en les associant à un nombre limité de mots de code (micro-états) et en contraignant la succession des mots de code choisis à préserver la stabilité temporelle observée. Les principales approches développées pour l’extraction des micro-états considèrent pour cela des algorithmes de partitionnement de données [9, 8]. Dans cette étude, nous proposons une généralisation du modèle des micro-états à l’aide d’une représentation redondante. L’extraction de ces états est alors réalisée grâce à un algorithme d’apprentissage de dictionnaire [10, 11], étant une généralisation de la quantification vectorielle. Au lieu de représenter un signal avec le mot de code le plus proche, celui-ci est représenté par une combinaison linéaire de quelques éléments du dictionnaire (appelés atomes).

Dans ce papier, nous étendons le modèle des micro-états à l’aide d’une représentation redondante dont le dictionnaire est obtenu par un algorithme d’apprentissage. L’approche considérée présente trois différences majeures avec l’approche des micro-états : (i) une contrainte de parcimonie moins stricte, (ii) l’intégration de la contrainte de stabilité temporelle dans l’étape de décomposition via une régularisation de type Variation Totale (VT) et (iii) la possibilité d’un dictionnaire contenant plus d’éléments que la dimension des topographies.

2 Modèle des N -micro-états

¹ Soit un enregistrement EEG $Y \in \mathbb{R}^{C \times T}$ mesuré à l'aide de C électrodes sur T échantillons temporels et $\Phi \in \mathbb{R}^{C \times N}$ une matrice non-redondante ($N \geq C$) rassemblant les topographies normalisées ($\forall n \in \mathbb{N}_N, \|\Phi(n)\|_2^2 = 1$) représentatives des micro-états. Le modèle des N -micro-états considère la décomposition linéaire du signal Y sur Φ suivante :

$$Y = \Phi X + E \quad \text{t.q.} \quad \forall t \in \mathbb{N}_T, \|X(t)\|_0 = 1, \quad (1)$$

avec E une matrice de bruit et X une matrice de coefficients dont les colonnes ne comporte qu'un seul élément non nul, *i.e.* $\forall t \in \mathbb{N}_T, \|X(t)\|_0 = 1$ (un seul état actif).

Ces micro-états sont appris sur un ensemble de M signaux $\{Y^m, \forall m \in \mathbb{N}_M\}$ par résolution du problème d'apprentissage (ME) qui suit :

$$\begin{aligned} (\{\hat{X}^m\}_{m=1}^M, \hat{\Phi}) &= \arg \min_{\{X^m\}_{m=1}^M, \Phi} \sum_{m=1}^M \|Y^m - \Phi X^m\|_F^2 \quad (2) \\ \text{t.q.} \quad \forall m \in \mathbb{N}_M, \forall t \in \mathbb{N}_T, \|X^m(t)\|_0 &= 1, \\ \text{et} \quad \forall n \in \mathbb{N}_N, \|\Phi(n)\|_2^2 &= 1. \end{aligned}$$

Plusieurs algorithmes de partitionnement ont été proposés pour la résolution de celui-ci : l'étude présentée dans [9] suggère une variante de l'algorithme des K-moyennes tandis qu'un algorithme de partitionnement hiérarchique est considéré dans [8]. L'apprentissage classique des micro-états s'effectue en alternant deux étapes [9]. À l'itération j , les micro-états les plus corrélés aux topographies sont tout d'abord sélectionnés. Cette opération s'effectue en maximisant la corrélation :

$$\tilde{n}(t) = \arg \max_n (\Phi_j^T(n) Y^m(t))^2. \quad (3)$$

Ensuite les micro-états sont mis à jour par décomposition en valeurs propres :

$$\Phi_{j+1}(n) = \arg \max_{\phi} \frac{\phi^T S_j^n \phi}{\phi^T \phi} \quad \text{avec} \quad S_j^n = \sum_{m=1}^M \sum_{t|\tilde{n}(t)=n} Y^m(t) Y^m(t)^T.$$

Ces deux étapes peuvent être vues comme un cas particulier d'apprentissage de dictionnaires [10], à sélection mono-atome.

En général, la stabilité temporelle des topographies n'est pas prise en compte directement : elle est assurée après une première sélection des micro-états, l'association de ces états avec les topographies étant pénalisée. Pour une topographie indiquée par t et un atome n , le terme de pénalité de lissage temporel est défini comme le nombre de topographies ayant été associées à n et dont les indices se situent à une distance temporelle inférieure à b de t , avec b un paramètre à déterminer :

$$\Theta^b(n, t) = \#\{\tilde{n}(s) = n \mid s \in [t - b, \dots, t + b]\}. \quad (4)$$

Ce terme est considéré dans une étape intermédiaire, donnant un algorithme d'apprentissage ME en trois étapes [9].

1. **Notation** : $\mathbb{N}_T = \{1, \dots, T\}$, $A(n)$ est la $n^{\text{ième}}$ colonne de la matrice A et $\|A\|_F$ sa norme de Frobenius.

3 Généralisation des N -micro-états

Dans le modèle des micro-états, chaque topographie d'un signal EEG est associée à un état et l'évolution de ces états est une fonction constante par morceaux. L'extraction des micro-états est donc un problème de QV. L'apprentissage de dictionnaires en est une généralisation : au lieu de représenter un signal avec le mot de code le plus proche, il est représenté comme la somme de quelques éléments du dictionnaire [10]. Nous proposons donc ici d'étendre le modèle des micro-états à l'aide d'une représentation redondante, dont le dictionnaire est obtenu via un algorithme d'apprentissage.

Le modèle considéré présente trois différences avec le modèle des micro-états : (i) la contrainte de parcimonie est moins stricte, *i.e.* $\|X(t)\|_0 \ll C$, autorisant plusieurs micro-états à être actifs simultanément, (ii) les coefficients de décomposition sont contraints à respecter l'*a priori* de stabilité temporelle des micro-états, permettant l'obtention de représentations plausibles physiologiquement, (iii) le dictionnaire peut contenir un nombre de micro-états plus important que la dimension des topographies, *i.e.* $N \geq C$.

Ces deux dernières caractéristiques sont imposées sur les matrices de décomposition grâce à une régularisation temporelle $R(X)$, combinant un terme de parcimonie relaxé en norme ℓ_1 $\|X\|_1$ et une pénalisation de type Variation Totale (VT) $\|XP\|_1$:

$$R(X) = \lambda \|X\|_1 + \mu \|XP\|_1, \quad (5)$$

avec P une matrice de différence finie encodant la régularisation VT sur la matrice X , ligne par ligne [12].

Le problème d'apprentissage de dictionnaires des micro-états généralisé (GenME) peut alors s'exprimer comme suit :

$$\begin{aligned} (\{\hat{X}^m\}_{m=1}^M, \hat{\Phi}) &= \arg \min_{\{X^m\}_{m=1}^M, \Phi} \sum_{m=1}^M \|Y^m - \Phi X^m\|_F^2 + R(X^m) \\ \text{et} \quad \forall n \in \mathbb{N}_N, \|\Phi(n)\|_2^2 &\leq 1. \end{aligned} \quad (6)$$

Afin de résoudre ce problème non-convexe, nous nous inspirons de l'algorithme d'apprentissage de dictionnaires en ligne introduit dans [11], dont chaque itération est composée de deux étapes, résolvant chacune un sous-problème convexe. Premièrement, la décomposition régularisée d'un $m^{\text{ième}}$ signal, choisi au hasard, s'écrit (à l'itération j) :

$$X_{j+1}^m = \arg \min_X \|Y^m - \Phi_j X\|_F^2 + \lambda \|X\|_1 + \mu \|XP\|_1,$$

et est résolue via l'algorithme présenté dans [12]. Le dictionnaire est ensuite mis à jour en fonction de la décomposition obtenue et de celles calculées dans les itérations précédentes par résolution du sous-problème suivant ² :

$$\Phi_{j+1} = \arg \min_{\Phi} \sum_{m=1}^j \|Y^m - \Phi X_{j+1}^m\|_F^2 \quad (7)$$

$$\text{t.q.} \quad \forall n \in \mathbb{N}_N, \|\Phi(n)\|_2^2 \leq 1.$$

et sa résolution est réalisée grâce à la mise à jour décrite dans [11].

2. Le relâchement de la contrainte sur la norme des atomes (par rapport au modèle ME) permet de conserver la convexité de ce sous-problème.

4 Résultats expérimentaux

Afin d'évaluer le modèle proposé, nous comparons les approches ME et GenME via deux expériences : la première sur données de synthèse et la seconde sur données réelles à travers l'étude du potentiel évoqué P300. L'algorithme de la Section 2 est adapté pour notre contexte redondant : même mise à jour que le GenME et lissage à chaque itération.

Apprentissage sur données de synthèse. Nous souhaitons comparer la capacité des deux approches à retrouver, à partir d'un ensemble de signaux présentant des structures sous-jacentes constantes par morceau, le dictionnaire utilisé pour les construire. Les structures considérées peuvent présenter des superpositions d'atomes, ce qui correspond à l'hypothèse de superposition des micro-états du modèle proposé. Afin de construire de tels signaux, un dictionnaire de synthèse Ψ est créé. Il est composé par $N = 30$ atomes de taille $C = 20$ tirés d'une distribution gaussienne (faible cohérence). Chaque signal de synthèse est ensuite conçu à partir du produit de ce dictionnaire et d'une matrice de coefficients constants par morceaux, avec $T = 300$. Ces matrices de coefficients sont construites comme sommes de K activités, définies par la présence d'un atome pendant une durée d , dont les superpositions sont aléatoires (voir [12]). L'expérience est réalisée pour différentes valeurs de K et de d .

Les modèles sont comparés en appliquant les algorithmes d'apprentissage ME et GenME sur ces signaux. Les dictionnaires appris, Φ_{ME} et Φ_{GenME} , possèdent le même nombre d'atomes N que le dictionnaire Ψ utilisé pour la création des signaux. Chacun des deux dictionnaires appris est alors comparé avec le dictionnaire Ψ , via la distance de Wasserstein sur les distances Euclidiennes entre atomes [13].

Pour chaque cas étudié (nombre K et durée d des activités), les paramètres de régularisation des algorithmes de décomposition sont déterminés sur $M = 100$ signaux d'entraînement. L'évaluation est ensuite réalisée sur $Q = 100$ signaux tests. De plus, du fait de la non-convexité du problème, les résultats sont moyennés sur 20 apprentissages.

Les résultats de cette expérience sont présentés dans la Fig. 1. Dans l'ensemble des cas observés, l'approche proposée présente de meilleures performances que l'algorithme d'extraction des micro-états. Le relâchement de la contrainte de parcimonie permet donc d'améliorer l'apprentissage du dictionnaire lorsque les structures de décomposition sous-jacentes présentent effectivement plusieurs atomes actifs simultanément.

Dans le cadre des signaux EEG, il semble raisonnable de supposer que les activités cérébrales ne sont pas disjointes temporellement. Notre approche semble donc adaptée à l'apprentissage des composantes spatiales sous-jacentes à ces signaux et nous allons donc l'appliquer à des signaux EEG réels.

Décomposition du potentiel évoqué P300. Les modèles sont maintenant utilisés pour analyser le P300. Dans un paradigme *oddball*, les stimuli visuels cibles engendrent un P300 contrai-

ME/GenME	Durée des activités d (en sec.)		
	$d = 0.3$	$d = 0.5$	$d = 0.7$
Nb act $K = 20$	0.4077/0.2184	0.5548/0.4188	0.7156/0.5300
Nb act $K = 30$	0.4634/0.3134	0.6449/0.4982	0.7298/0.5917
Nb act $K = 40$	0.5867/0.3979	0.6713/0.5754	0.7903/0.5863
Nb act $K = 50$	0.5838/0.4619	0.7401/0.5746	0.8063/0.6189

FIGURE 1 – Comparaison des distances entre les dictionnaires appris et le dictionnaire de synthèse : modèle des micro-états ME (rouge) *versus* modèle généralisé GenME (bleu).

rement aux stimuli non-cibles [14]. Sans « vérité terrain » concernant les composantes sous-jacentes du P300, il est difficile d'évaluer les résultats obtenus. Par conséquent, les décompositions obtenues avec l'extension GenME sont comparées à celles obtenues avec le modèle classique ME du point de vue de la capacité de description.

Nous considérons ici les signaux Iib de la 2^{ème} compétition BCI [15], mesurés sur $C = 64$ électrodes à 240Hz, et enregistrés sur un seul sujet au cours de 3 sessions de mesure. Les essais sont extraits dans l'intervalle 0-600 ms après chaque stimulus et filtrés entre 0.3 et 20Hz. Les paramètres de régularisation des deux approches sont appris sur les essais (non-moyennés) en sélectionnant ceux permettant d'obtenir des décompositions dont les principales ruptures (maxima du terme VT), présentent des distances temporelles minimales par rapport aux minima du *Global Field Power*³ (GFP)[8]. Les moyennes des essais sont calculées pour chaque session pour les deux classes : réponses P300 et réponses non-cibles. Les dictionnaires sont appris sur les $M = 576$ topographies des signaux moyens des deux premières sessions (144 échantillons \times 2 PE \times 2 sessions), puis évalués en décomposant les signaux moyens de la troisième.

Les représentations obtenues pour les signaux moyens des deux classes par les deux approches sont présentées dans la Fig. 2 pour les atomes les plus représentatifs. Ces représentations sont mises en regard avec le GFP des topographies. Les deux approches font apparaître les mêmes ruptures dans les signaux et les zones de stabilité correspondent aux instants où l'indice GFP est maximal. Concernant le modèle classique ME, les micro-états extraits ressemblent fortement à ceux de la littérature [16]. De plus, la séparation en deux parties du P300 est une caractéristique connue de ce dernier qui a notamment été décrite dans [9]. Quant à elle, la décomposition GenME comporte plus de ruptures et d'activités de faible durée que celle du modèle classique. Prenant en compte la possibilité de superposition des activités, cette généralisation permet la détection de ruptures entre les activités avec plus de finesse, même lorsque les topographies de celles-ci sont similaires. Cela permet notamment de capturer des changements d'activités n'apparaissant pas dans les évolutions de l'indice GFP. Dans cette décomposition, nous observons que le P300 est composé de 4 activités. Les deux atomes les plus marqués (coefficients plus forts pour le 4 et le 6) ressemblent fortement aux atomes extraits dans le modèle classique, mais deux autres atomes sont activés également avec des coefficients plus faibles. Les intervalles

3. GFP : $P_y(t) = \sqrt{\frac{1}{C} \sum_{c=1}^C (y(c,t) - \bar{y}_c)^2}$ et $\bar{y}_c = \frac{1}{T} \sum_{t=1}^T y(c,t)$

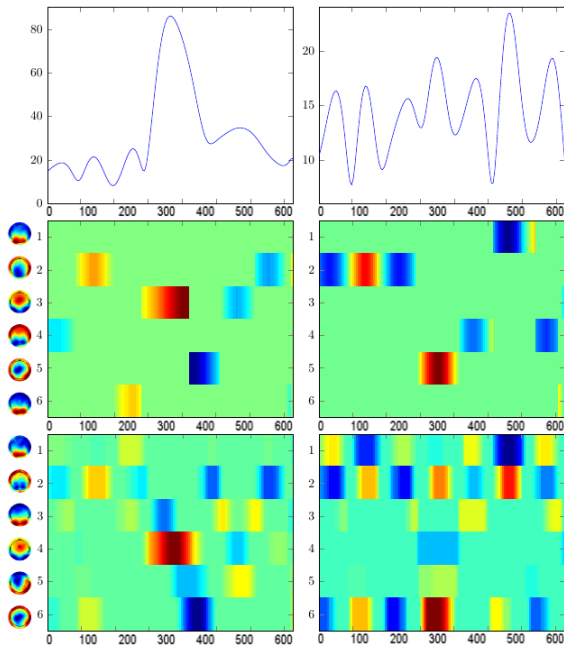


FIGURE 2 – Représentation de la réponse P300 (gauche) et de la réponse non-cible (droite), avec pour chaque colonne, l’indice GFP (haut), la décomposition et les atomes topographiques associés, pour le modèle ME (centre) et pour le modèle GenME (bas). Le temps est exprimé en millisecondes.

temporels d’activités de ces atomes n’étant pas disjoints, le modèle classique ne peut les faire apparaître. Le modèle GenME est aussi capable de capturer la réponse aux stimulations lumineuses à 5.7Hz [14]. La fréquence de la puissance est double, et est bien visible grâce à l’atome 2, autant sur la réponse cible que non-cible [14]. Cela n’est pas le cas sur le modèle ME.

5 Conclusion

Une généralisation du modèle des micro-états considérant une représentation redondante est proposée dans ce travail. L’extraction de ces états est réalisée à l’aide d’un algorithme d’apprentissage de dictionnaires dont l’étape de décomposition est contrainte à l’aide d’une régularisation de type fused-LASSO multi-dimensionnel afin de prendre en compte la stabilité temporelle des micro-états. Les expériences réalisées montrent d’une part que l’approche proposée permet de retrouver plus efficacement les composantes sous-jacentes lorsque les signaux respectent les hypothèses de stabilité et que plusieurs d’entre elles sont simultanément actives et d’autre part qu’elle permet une description plus fine du potentiel évoqué P300.

De futurs travaux pourront considérer d’une part la création de méthodes permettant de déterminer les hyperparamètres de cette approche et d’autre part son application à la décomposition d’autres types de signaux EEG.

Remerciements Ces travaux s’inscrivent dans le cadre du projet Unsupervised-Brain (2011-053D) financé par DIGITEO.

Références

- [1] D Lehmann, H Ozaki, and I Pal. EEG alpha map series : brain micro-states by space-oriented adaptive segmentation. *Electroencephalogr Clin Neurophysiol*, 67 :271–288, 1987.
- [2] D Lehmann. Brain electric microstates and cognition : the atoms of thought. In *Machinery of the Mind*, pages 209–224. 1990.
- [3] D Lehmann, PL Faber, S Galderisi, WM Herrmann, T Kinoshita, M Koukkou, A Mucci, RD Pascual-Marqui, et al. EEG microstate duration and syntax in acute, medication-naive, first-episode schizophrenia : a multi-center study. *Psychiatry Res*, 138 :141–156, 2005.
- [4] WK Strik, T Dierks, T Becker, and D Lehmann. Larger topographical variance and decreased duration of brain electric microstates in depression. *J Neural Transm Gen Sect*, 99 :213–222, 1995.
- [5] H Yuan, V Zotev, R Phillips, WC Drevets, and J Bodurka. Spatiotemporal dynamics of the brain at rest-exploring EEG microstates as electrophysiological signatures of BOLD resting state networks. *Neuroimage*, 60 :2062–2072, 2012.
- [6] AJ Pegna, A Khateb, L Spinelli, M Seeck, T Landis, and CM Michel. Unraveling the cerebral dynamics of mental imagery. *Hum Brain Mapp*, 5 :410–421, 1997.
- [7] M Wirth, H Horn, T Koenig, A Razafimandimby, M Stein, T Mueller, A Federspiel, B Meier, et al. The early context effect reflects activity in the temporo-prefrontal semantic system : evidence from electrical neuroimaging of abstract and concrete word reading. *NeuroImage*, 42 :423–436, 2008.
- [8] MM Murray, D Brunet, and CM Michel. Topographic ERP analyses : a step-by-step tutorial review. *Brain topography*, 20 :249–264, 2008.
- [9] RD Pascual-Marqui, CM Michel, and D Lehmann. Segmentation of brain electrical activity into microstates : model estimation and validation. *IEEE Trans Biomed Eng*, 42 :658–665, 1995.
- [10] I. Tošić and P. Frossard. Dictionary learning. *IEEE Signal Process. Mag.*, 28 :27–38, 2011.
- [11] J Mairal, F Bach, J Ponce, and G Sapiro. Online dictionary learning for sparse coding. In *ICML*, pages 689–696, 2009.
- [12] Y Isaac, Q Barthélemy, J Atif, C Gouy-Pailler, and M Sebag. Multi-dimensional sparse structured signal approximation using split Bregman iterations. In *ICASSP*, pages 3826–3830, 2013.
- [13] S Chevallier, Q Barthélemy, and J Atif. On the need for metrics in dictionary learning assessment. In *EUSIPCO*, pages 1427–1431, 2014.
- [14] B. Rivet, A. Souloumiac, G. Gibert, and V. Attina. "P300 speller" brain-computer interface : Enhancement of P300 evoked potential by spatial filters. In *EUSIPCO*, 2008.
- [15] B Blankertz, K Müller, G Curio, TM Vaughan, G Schalk, JR Wolpaw, A Schlögl, C Neuper, G Pfurtscheller, et al. The BCI competition 2003 : progress and perspectives in detection and discrimination of EEG single trials. *IEEE Trans Biomed Eng*, 51 :1044–1051, 2004.
- [16] D Lehmann, RD Pascual-Marqui, and C Michel. EEG microstates, www.scholarpedia.org/article/eeg_microstates, 2013.