

Apprentissage de dictionnaire convolutif multivarié pour les signaux EEG et MEG

Thomas MOREAU¹, Tom DUPRÉ LA TOUR², Mainak JAS³, Alexandre GRAMFORT¹

¹Parietal, INRIA Saclay, Université Paris-Saclay, 1 rue Honoré Estienne d’Orves, 91120 Saclay, France

²Helen Wills Neuroscience Institute, University of California, Berkeley

³Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, USA

thomas.moreau@inria.fr, tomldt@berkeley.edu

mjas@mgh.harvard.edu, alexandre.gramfort@inria.fr

Résumé – Dans cet article, nous proposons d’apprendre des représentations dédiées aux signaux d’électroencéphalographie (EEG) et magnétoencéphalographie (MEG) en utilisant un algorithme multivarié de codage convolutif parcimonieux (convolutional sparse coding, CSC) avec des contraintes de rang 1. Notre méthode est capable d’apprendre non seulement des formes d’onde temporelles prototypiques, mais également les modèles spatiaux associés afin que leur origine puisse être localisée dans le cerveau. L’algorithme d’estimation est basé sur une minimisation alternée et une méthode localement gloutonne de descente par coordonnées offrant des performances améliorées par rapport à l’état de l’art.

Abstract – In this paper, we propose to learn dedicated representations of electroencephalography (EEG) and magnetoencephalography (MEG) signals using a multivariate convolutional sparse coding (CSC) algorithm with rank-1 constraints. Applied to electrophysiological signals, this method is able to learn not only prototypical temporal waveforms, but also associated spatial patterns so their origin can be localized in the brain. Our algorithm is based on alternated minimization and a locally greedy coordinate descent solver that leads to state-of-the-art performances.

1 Introduction

L’activité neuronale enregistrée via des mesures du potentiel électrique à la surface de la peau par électroencéphalographie (EEG) ou des champs magnétiques par magnétoencéphalographie (MEG) peut être utilisée pour étudier les processus cognitifs humains ainsi que certaines pathologies. Ces enregistrements sont constitués de dizaines ou de centaines de signaux enregistrés simultanément, pour des durées allant de quelques minutes à quelques heures. Pour étudier les événements transitoires ainsi que la morphologie des formes d’onde, éléments critiques pour la cognition et pour notre compréhension des pathologies [1], il est nécessaire d’avoir recours à des techniques différant du traitement du signal traditionnel. Par exemple, une analyse classique de Fourier ne permet pas de distinguer les rythmes alpha des rythmes mu, qui ont la même fréquence principale autour de 10 Hz, mais dont les formes d’onde diffèrent [2].

L’estimation de représentations à partir des données est la clé de nombreuses techniques d’analyses statistiques modernes appliquées à des données complexes telles que des images naturelles, des sons ou des séries temporelles biologiques. L’apprentissage par dictionnaire est une famille de techniques consistant à apprendre des motifs (ou atomes) permettant de reconstruire approximativement les données de manière parcimonieuse. Lorsque l’on travaille avec des signaux longs dans lesquels des événements peuvent survenir à tout moment, une

approche est d’apprendre des atomes *invariant par translation*. Ils peuvent alors offrir de meilleures approximations de signaux que les bases génériques telles que Fourier ou les ondelettes, car ils ne se limitent pas à des bandes de fréquences étroites.

Une approche populaire pour l’apprentissage de dictionnaire invariant par translation est le modèle de codage parcimonieux convolutif (CSC) [3]. L’idée est de poser le problème comme un problème d’optimisation, représentant le signal comme une somme de convolutions entre atomes et signaux d’activation. Dans le cas des enregistrements EEG et MEG, l’activité neuronale est instantanément et linéairement répartie sur tous les canaux, d’après les équations de Maxwell [4]. Les mêmes modèles temporels sont donc reproduits sur tous les canaux avec des intensités différentes qui dépendent de l’emplacement de chaque activité dans le cerveau. Pour exploiter cette propriété, nous proposons d’utiliser une contrainte de rang 1 sur chaque atome. Cette idée a déjà été mentionnée par Barthélemy et al. [5], mais a été jugée moins souple que le modèle de rang complet. De plus, la techniques d’optimisation proposée dans [5] n’est pas spécifique aux modèles invariants par translation, et n’est pas adaptée aux signaux longs. Une telle décomposition EEG invariante par translation et de rang 1 a également été proposée à l’aide de poursuite gloutonne [6], mais sans apprendre les atomes, qui sont alors fixés selon des filtres de Gabor.

Contribution Dans cette étude, nous développons un modèle multivarié pour CSC, utilisant une contrainte de rang 1

sur les atomes pour rendre compte de la propagation instantanée d'une source électromagnétique sur tous les canaux. Nous proposons également des stratégies d'optimisation efficaces, à savoir une descente par coordonnées localement gloutonne (LGCD) [7], ainsi que des étapes de précalcul pour accélérer les calculs de gradients. Nous fournissons plusieurs évaluations numériques de notre méthode, qui montrent un temps d'exécution très compétitif sur les modèles univariés et multivariés, même avec des centaines de canaux. Enfin, nous illustrons notre méthode avec des atomes appris sur des données multivariées MEG, qui, grâce au modèle de rang 1, peuvent être localisés dans le cerveau pour des études de neurosciences cliniques ou cognitives. Ce travail est disponible plus en détail dans Dupré la Tour et al. [8].

Notation On note $[a, b]$ (*resp.* $\llbracket a, b \rrbracket$) l'ensemble des réels (*resp.* entiers) entre a et b . Un signal multivarié avec T pas de temps dans \mathbb{R}^P est noté $X \in \mathbb{R}^{P \times T}$, et on note $x \in \mathbb{R}^T$ les signaux univariés. Le temps est indexé avec des crochets $X[t] \in \mathbb{R}^P$, alors que $X_i \in \mathbb{R}^T$ correspond au canal i de X . Pour un signal multivarié $X \in \mathbb{R}^{P \times T}$, la norme ℓ_p est définie temporellement par $\|X\|_q = (\sum_{t=1}^T \|X[t]\|_q^q)^{1/q}$. On note U^\top la transposée de la matrice U . Pour un signal multivarié $X \in \mathbb{R}^{P \times T}$, X^\natural est le signal obtenu en renversant le temps dans X , *i.e.* $X^\natural[t] = X[T+1-t]^\top$. On notera \tilde{T} pour $T-L+1$.

2 Dictionnaire convolutif appris

La formulation du CSC adoptée dans ce papier suit le modèle de codage parcimonieux invariant par translation introduit par Grosse et al. [3]. Pour un jeu de N signaux multivarié $X^n \in \mathbb{R}^{P \times T}$, on cherche à apprendre une famille de K atomes $D_k \in \mathbb{R}^{P \times L}$ ainsi que les K activations qui leur sont associées pour chaque signal $z_k^n \in \mathbb{R}^{\tilde{T}}$. Le modèle que nous proposons est similaire au modèle classique, auquel on ajoute une contrainte de rang-1 pour les atomes du dictionnaire, $D_k = u_k v_k^\top \in \mathbb{R}^{P \times L}$, où $u_k \in \mathbb{R}^P$ représente le motif sur les canaux et $v_k \in \mathbb{R}^L$ le motif temporel. L'estimation des paramètres revient alors à :

$$\min_{u_k, v_k, z_k^n} \sum_{n=1}^N \frac{1}{2} \left\| X^n - \sum_{k=1}^K z_k^n * (u_k v_k^\top) \right\|_2^2 + \lambda \sum_{k=1}^K \|z_k^n\|_1, \quad (1)$$

t.q. $\|u_k\|_2^2 \leq 1$, $\|v_k\|_2^2 \leq 1$ et $z_k^n \geq 0$.

La contrainte de rang est cohérente avec le modèle physique de l'activité électrophysiologique EEG et MEG régi par les équations de Maxwell. En effet, les fonctions cognitives sont associées à une activité neuronale locale (et localisée) dans le cerveau. Pour ces activités, chaque source est propagée linéairement et instantanément sur les canaux [4]. De plus, l'utilisation de cette hypothèse peut améliorer l'estimation de motifs en présence de bruit indépendant sur chaque canal, et aussi mieux séparer les sources qui se chevauchent. Ce modèle est aussi plus efficace en terme de calcul car certaines opérations peuvent être factorisées. La Figure 1 résume ce modèle.

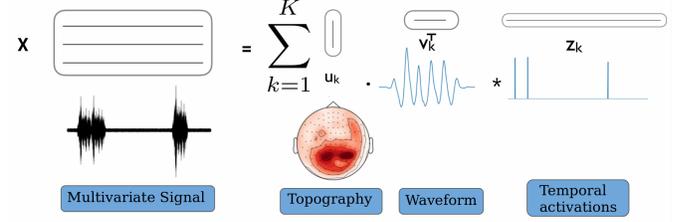


FIGURE 1 – Illustration du modèle de codage convolutif parcimonieux de rang 1.

3 Estimation et optimisation

Le problème (1) est tri-convexe, *i.e.* il est convexe pour chaque variable, mais pas conjointement convexe. Il n'y a donc pas garantie d'identifiabilité de ce modèle. Une solution approchée est obtenue en cherchant un minimum local du problème (1). L'algorithme proposé est basé sur une descente par coordonnées en blocs, qui minimise alternativement la fonction objectif sur chaque bloc de variables.

3.1 Étape Z : calcul des activations

Étant donné K atomes *fixés* D_k et un paramètre de régularisation $\lambda > 0$, l'étape Z vise à calculer pour chacun des N signaux $X^n \in \mathbb{R}^{P \times T}$ les K signaux d'activation $z_k^n \in \mathbb{R}^{\tilde{T}}$ associés en résolvant le problème d'optimisation suivant :

$$\min_{z_k^n \geq 0} \frac{1}{2} \left\| X^n - \sum_{k=1}^K z_k^n * D_k \right\|_2^2 + \lambda \sum_{k=1}^K \|z_k^n\|_1. \quad (2)$$

Ce problème est indépendant pour chaque signal X^n . Le calcul de chaque z^n peut donc être parallélisé, indépendamment de la technique choisie pour résoudre l'optimisation [9]. Par conséquent, nous omettons l'exposant n dans la sous-section suivante pour simplifier la notation. Le problème (2) est convexe en z_k et peut être résolu efficacement par méthodes itératives. Ici, nous utilisons la descente par coordonnées localement gloutonne (LGCD) proposée récemment par Moreau et al. [7], particulièrement efficace pour les signaux longs et parcimonieux.

Descente par coordonnée localement gloutonne L'idée de la descente par coordonnées est de mettre à jour l'estimé de la solution une coordonnée $z_k[t]$ à la fois. Pour (2), il est possible de calculer la valeur optimale $z_k'[t]$ d'une coordonnée $z_k[t]$ lorsque toutes les autres sont fixes. En effet, le problème (2) limité à une coordonnée a une expression analytique :

$$z_k'[t] = \max \left(\frac{\beta_k[t] - \lambda}{\|D_k\|_2^2}, 0 \right), \quad \text{avec} \quad (3)$$

$$\beta_k[t] = \left[D_k^\natural \tilde{*} \left(X - \sum_{l=1}^K z_l * D_l + z_k[t] e_t * D_k \right) \right] [t]$$

où $e_t \in \mathbb{R}^{\tilde{T}}$ est le vecteur canonique qui a pour valeur 1 à l'index t et 0 ailleurs. Lorsque le coefficient $z_{k_0}[t_0]$ est mis à

Algorithm 1: Descente par coordonnée localement gloutonne (LGCD)

Entrée: Signal X , atomes D_k , nombre de segments M , paramètre d'arrêt $\epsilon > 0$, z_k initialisation

Initialiser $\beta_k[t]$ avec (3).

repeat

for $m = 1$ **to** M **do**

Pour $(k, t) \in \mathcal{C}_m$; $z'_k[t] = \max\left(\frac{\beta_k[t] - \lambda}{\|D_k\|_2^2}, 0\right)$

Choisir $(k_0, t_0) = \arg \max_{(k,t) \in \mathcal{C}_m} |z_k[t] - z'_k[t]|$

Mettre à jour β avec (4)

Mettre à jour l'estimé courante $z_{k_0}[t_0] \leftarrow z'_{k_0}[t_0]$

until $\|z - z'\|_\infty < \epsilon$;

TABLE 1 – Complexités des opérations

Étape : Calcul	Rang-1	Rang plein
Z-step : Initial $\beta^{(0)}$	$NKT(L + P)$	$NKT(LP)$
Z-step : Constante $D^\dagger * D$	$K^2L(L + P)$	$K^2L(LP)$
Z-step : M mise à jours	MKL	MKL
D-step : Constante Φ	$NKTLP$	$NKTLP$
D-step : Constante Ψ	NK^2TL	NK^2TLP
D-step : Gradient $\nabla_{D_k} E$	$K^2L(L + P)$	$K^2L(LP)$
D-step : Coût E	$K^2L(L + P)$	$K^2L(LP)$

jour à la valeur $z'_{k_0}[t_0]$, β change pour $(k, t) \neq (k_0, t_0)$ de

$$\beta_k^{(q+1)}[t] = \beta_k^{(q)}[t] + (D_{k_0}^\dagger \tilde{*} D_k)[t - t_0](z_{k_0}[t_0] - z'_{k_0}[t_0]). \quad (4)$$

Le terme $(D_{D_{k_0}}^\dagger \tilde{*} D_k)[t - t_0]$ est nul pour $|t - t_0| \geq L$. Ainsi, seuls $K(2L - 1)$ coefficients de β doivent être modifiés [10]. L'algorithme met à jour à chaque itération une seule coordonnée à cette valeur optimale. L'idée proposée dans [7] est de sous diviser les coordonnées en M segments et de choisir à l'itération q la coordonnée (k_0, t_0) qui est la plus loin de sa valeur optimale sur un sous ensemble contigu de coordonnées $\mathcal{C}_m = \llbracket 1, K \rrbracket \times \llbracket (m-1)\tilde{T}/M, m\tilde{T}/M \rrbracket$ i.e.

$$(k_0, t_0) = \arg \max_{(k,t) \in \mathcal{C}_m} |z_k[t] - z'_k[t]|, \quad m \equiv q \pmod{M} + 1. \quad (5)$$

Avec cette stratégie, la complexité de la sélection des coordonnées est linéaire dans la longueur du sous-segment considéré $\mathcal{O}(K\tilde{T}/M)$. En choisissant $M = \lfloor \tilde{T}/(2L - 1) \rfloor$, la complexité de la mise à jour est équivalente à celle des sélections aléatoire et cyclique, $\mathcal{O}(KL)$. Cependant, cette technique permet de privilégier les coordonnées plus importante par rapport à une sélection aléatoire de la coordonnée, améliorant sa convergence. Nous détaillons les étapes de LGCD dans l'Algorithme 1.

3.2 Étape D : calcul des atomes

Étant donné KN signaux d'activation fixes $z_k^n \in \mathbb{R}^{\tilde{T}}$ associés aux signaux $X^n \in \mathbb{R}^{P \times T}$, l'étape D vise à mettre à jour

K motifs spatiaux $u_k \in \mathbb{R}^P$ et K motifs temporels $v_k \in \mathbb{R}^L$, par la résolution du problème suivant

$$\min_{\substack{\|u_k\|_2 \leq 1 \\ \|v_k\|_2 \leq 1}} E(u_k, v_k) \triangleq \sum_{n=1}^N \frac{1}{2} \|X^n - \sum_{k=1}^K z_k^n * (u_k v_k^\top)\|_2^2. \quad (6)$$

Le problème (6) est convexe dans chaque bloc de variables $\{u_k\}_k$ et $\{v_k\}_k$, mais pas conjointement. Par conséquent, nous optimisons d'abord $\{u_k\}_k$, puis $\{v_k\}_k$, en utilisant dans les deux cas une descente de gradient projeté avec une recherche linéaire de type Armijo [11] pour la taille du pas de gradient.

Gradient relatif à u_k et v_k Le gradient de E par rapport à u_k et v_k peut être calculé en utilisant les règles des fonctions composées. L'opération critique pour ce étape est le calcul du gradient par rapport à un atome complet $D_k = u_k v_k^\top \in \mathbb{R}^{P \times L}$:

$$\begin{aligned} \nabla_{D_k} E(u_k, v_k) &= \sum_{n=1}^N (z_k^n)^\dagger * \left(X^n - \sum_{l=1}^K z_l^n * D_l \right) \\ &= \Phi_k - \sum_{l=1}^K \Psi_{k,l} * D_l, \end{aligned} \quad (7)$$

où l'on réordonne l'équation (7) pour définir les constantes $\Phi_k \in \mathbb{R}^{P \times L}$ et $\Psi_{k,l} \in \mathbb{R}^{2L-1}$. Celles-ci peuvent être pré-calculées pour accélérer le calcul du gradient et de la fonction de coût E . Les complexités de calcul sont détaillées dans le Tableau 1. On peut noter que la dépendance en N et T est présente *seulement* dans les pré-calculs, rendant ainsi les itérations très rapides. Sans ces pré-calculs, la complexité de *chaque* gradient dans l'étape-D est de $\mathcal{O}(NK^2TLP)$.

4 Expériences numériques

Les expériences ont toutes été réalisées en Python et notre code est accessible en ligne¹.

Enregistrement somatosensoriel Notre méthode est évaluée sur un jeu de données MEG obtenu lors d'une simulation somatosensorielles du nerf médian. Les données sont fournies par la librairie MNE [12]. Ces enregistrements sont rééchantillonnés à 150 Hz, i.e. $T = 134\,700$ et $N = 2$. Nous ne considérons que les canaux des gradiomètres, conduisant à $P = 204$. Les signaux ont été filtrés avec un filtre passe-bande pour éliminer le bruit de l'alimentation à 50 Hz et filtrés en passe-haut à 2 Hz pour éliminer la tendance des basses fréquences.

Performance de l'algorithme La performance de notre stratégie d'optimisation est évaluée sur ce jeu de données MEG. Nous avons utilisé les paramètres $K = 8$ et $L = 128$. Notre méthode est comparée à un algorithme de type ADMM multi-varié développé par Wohlberg [13]. Cet algorithme étant assez lent sur les signaux longs, et même très lent à mesure que P augmente, nous avons limité nos expériences à $P = 5$ canaux.

1. <https://alphacsc.github.io/>

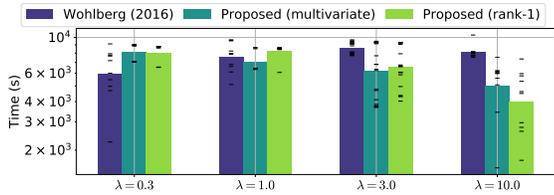


FIGURE 2 – Temps pour atteindre une précision relative de 10^{-3} , pour différents paramètre de régularisation λ avec $P = 5$.

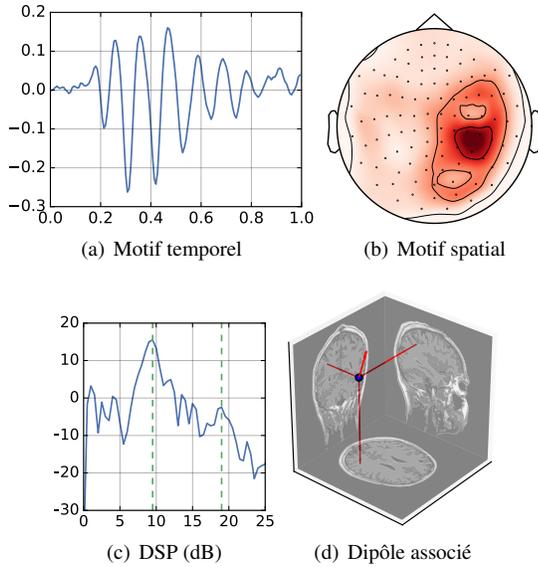


FIGURE 3 – Atome non sinusoidal manuellement sélectionné, appris dans les données somatosensorielles de MNE.

Les résultats, présentés dans Figure 2, montrent que notre méthode est plus rapide que la méthode concurrente pour les fortes valeurs de λ , qui est le régime le plus intéressant pour les applications considérées.

Exemples d’atomes obtenus sur signaux MEG Nous montrons les résultats de notre algorithme sur des données expérimentales, en utilisant le jeu de données somatosensorielles de MNE décrit ci-dessus. Nous avons appris $K = 40$ atomes avec $L = 150$ en utilisant un modèle CSC multivarié de rang 1, avec une régularisation $\lambda = 0.2\lambda_{max}$, où λ_{max} est la plus petite valeur de λ pour laquelle 0 est solution de (2) avec le dictionnaire initial. La régularisation est choisie de manière manuelle par rapport à la qualité des atomes produits. L’automatisation de ce choix est un problème ouvert du fait du caractère non supervisé de cette méthode.

Figure 3(a) montre un rythme cérébral non sinusoidal connu sous le nom de rythme mu en raison de sa forme ressemblant à la lettre grecque. Le rythme mu est impliqué dans l’activité motrice [14] et est centré autour de 10 Hz. Bien que la puissance soit concentrée dans la même bande de fréquences que le rythme alpha, il a une topographie spatiale très différente (Figure 3(b)). Dans Figure 3(c), la densité spectrale de puissance

(DSP) montre deux composantes du rythme mu - une à environ 9 Hz, et une harmonique à 18 Hz comme précédemment rapporté dans [14]. D’après notre analyse, il est clair que la composante 18 Hz est simplement une harmonique du rythme mu, mais une analyse basée sur Fourier pourrait nous amener à conclure à tort que les données contiennent des rythmes bêta. Enfin, en raison de la nature de rang 1 de nos atomes, il est possible d’ajuster un dipôle de courant équivalent [15] pour interpréter l’origine du signal. Le dipôle de courant équivalent (ECD) est calculé à partir d’une carte spatiale u_k et des équations de diffusion spécifique au sujet considéré. L’algorithme recherche la localisation dans le cerveau et l’orientation d’un dipôle qui génère la carte la plus proche de u_k . Figure 3(d) montre que l’atome se localise effectivement dans le cortex somatosensoriel primaire (S1).

Références

- [1] S. R. Jones. When brain rhythms aren’t ‘rhythmic’: implication for their mechanisms and meaning. *Curr. Opin. Neurobiol.*, 40 :72–80, 2016.
- [2] S. R. Cole and B. Voytek. Brain oscillations and the importance of waveform shape. *Trends Cogn. Sci.*, 2017.
- [3] R. Grosse, R. Raina, H. Kwong, and A. Y. Ng. Shift-invariant sparse coding for audio classification. In *23rd Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 149–158. AUAI Press, 2007. ISBN 0-9749039-3-0.
- [4] R. Hari and A. Puce. *MEG-EEG Primer*. Oxford University Press, 2017.
- [5] Q. Barthélemy, A. Larue, A. Mayoue, D. Mercier, and J. I. Mars. Shift & 2D rotation invariant sparse coding for multivariate signals. *IEEE Transactions on Signal Processing*, 60(4) :1597–1611, 2012.
- [6] P. J. Durka, A. Matysiak, E. M. Montes, P. V. Sosa, and K. J. Blinowska. Multichannel matching pursuit and EEG inverse solutions. *J. Neurosci. Methods*, 148(1) :49–59, 2005.
- [7] T. Moreau, L. Oudre, and N. Vayatis. DICOD : Distributed Convolutional Sparse Coding. In *International Conference on Machine Learning (ICML)*, 2018.
- [8] Tom Dupré la Tour, Thomas Moreau, Mainak Jas, and Alexandre Gramfort. Multivariate convolutional sparse coding for electromagnetic brain signals. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 3292–3302. Curran Associates, Inc., 2018.
- [9] M. Jas, T. Dupré La Tour, U. Şimşekli, and A. Gramfort. Learning the morphology of brain signals using alpha-stable convolutional sparse coding. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1–15, 2017.
- [10] K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, and Y. Le Cun. Learning convolutional feature hierarchies for visual recognition. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1090–1098, 2010.
- [11] S. Wright and J. Nocedal. *Numerical optimization*, volume 35. Springer Science, 1999.
- [12] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience*, 7, 2013.
- [13] B. Wohlberg. Convolutional sparse representation of color images. In *IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, pages 57–60, 2016.
- [14] R. Hari. Action-perception connection and the cortical mu rhythm. *Progress in brain research*, 159 :253–260, 2006.
- [15] T. Tuomisto, R. Hari, T. Katila, T. Poutanen, and T. Varpula. Studies of auditory evoked magnetic and electric responses : Modality specificity and modelling. *Il Nuovo Cimento D*, 2(2) :471–483, 1983.