

# Estimation régularisée et non supervisée de la réponse hémodynamique en imagerie cérébrale fonctionnelle (IRMf)

Philippe CIUCIU<sup>1</sup>, Jérôme IDIER<sup>2</sup>, Guillaume MARRELEC<sup>3</sup>, Jean-Baptiste POLINE<sup>1</sup>

<sup>1</sup>Service Hospitalier Frédéric Joliot (CEA/DSV/DRM)  
4 Place du Général Leclerc, 91401 Orsay cedex, France

<sup>2</sup>IRCCyN (CNRS/ECN/UdN/EMN)  
1 rue de la Noë, BP 92101 44321 Nantes cedex 3, France

<sup>3</sup>INSERM U 494 CHU Pitié-Salpêtrière  
91 bd de l'Hôpital, 75634 Paris cedex 13, France  
ciuciu@shfj.cea.fr, Jerome.Idier@irc cyn.ec-nantes.fr  
Guillaume.Marrelec@imed.jussieu.fr, poline@shfj.cea.fr

**Résumé** – L'estimation de la fonction de réponse hémodynamique (FRH) en imagerie par résonance magnétique fonctionnelle (IRMf) est essentielle pour une meilleure compréhension des activations cérébrales. Nous abordons ce problème dans un cadre bayésien en introduisant un *a priori* temporel sur la FRH et sous une forme *non supervisée* en maximisant la log-vraisemblance vis-à-vis des hyperparamètres. L'originalité de ce travail réside dans la définition d'une nouvelle fonction de vraisemblance, au nombre de paramètres réduit, qui vise d'une part à améliorer la prise en compte de la variabilité des artefacts physiologiques (cœur, respiration), et d'autre part à accélérer la convergence de l'algorithme EM utilisé pour la maximiser. Nous montrons l'intérêt de cette approche en la comparant aux travaux pré-existants [1], à la fois en simulation et sur données réelles.

**Abstract** – Hemodynamic Response Function (HRF) estimation in noisy functional Magnetic Resonance Imaging (fMRI) is essential for a better understanding of cerebral activations. We address this problem in the Bayesian framework and introduce temporal prior information on the HRF. The unsupervised feature of the HRF estimate relies on the maximization of the log-likelihood with respect to the hyperparameters. In this paper, we focus on the definition of a new likelihood function that accounts better for the variability of the physiological artifacts (breathing and cardiac rates). This function also depends on less parameters, so that faster convergence is obtained for the EM algorithm used as likelihood maximization. We demonstrate the relevance of this approach through a comparison with pre-existing works [1] both on synthetic and real data.

## 1 Introduction

L'IRMf a permis le développement des études d'activation cérébrale dans lesquelles on enregistre au cours du temps un paramètre *hémodynamique* lié à l'activité du cerveau [2]. Grâce à sa bonne résolution spatiale, l'IRMf permet de cartographier cette activité lorsque le sujet perçoit un stimulus ou réalise une tâche [3]. La faible résolution temporelle de l'IRMf l'empêche toutefois de suivre la dynamique des activations au niveau neuronal. En effet, le couplage entre l'activation des neurones et le système vasculaire, en sortie duquel le signal IRMf est mesuré, introduit un délai hémodynamique et un filtrage passe-bas qu'il convient de caractériser pour tenter d'augmenter la résolution temporelle. Beaucoup d'approches ont été proposées pour estimer ce filtre, appelé fonction de réponse hémodynamique (FRH). Celles développées dans [4, 1] sont non paramétriques mais relativement robustes car elles intègrent dans le cadre *bayésien* des informations *a priori* sur sa structure temporelle.

Nous présentons dans cette communication une alternative à la démarche retenue dans [1] pour estimer les paramètres du modèle et tenir compte des artefacts physiologiques présents dans les données IRMf. Nous définissons pour cela une nouvelle fonction de vraisemblance au nombre de paramètres

réduit, accélérant la convergence de l'algorithme Expectation Maximization (EM) utilisé pour la maximiser. L'estimateur de la FRH qui en découle traduit mieux la variabilité de ces artefacts. Nous démontrons l'intérêt de cette nouvelle approche tant en simulation que sur données réelles issues d'une expérience sur la compréhension du langage.

## 2 Modélisation de la FRH

### 2.1 Modèle simplifié

Le signal IRMf  $\mathbf{y} = [y_{t_1}, \dots, y_{t_N}]^t$ , acquis dans chaque *voxel* du cerveau (*e.g.*,  $3 \times 3 \times 3 \text{ mm}^3$ ), est mesuré aux temps ( $t_n = n \text{ TR}$ ) $_{1 \leq n \leq N}$ , où *TR* désigne le délai entre deux acquisitions. Sous l'hypothèse d'un couplage vasculaire cérébral *linéaire* et *invariant* dans le temps, la FRH désigne la réponse impulsionnelle inconnue  $\mathbf{h} = [h_0, h_{\Delta t}, \dots, h_{P\Delta t}]^t$  du système recevant en entrée la séquence binaire  $(\mathbf{x}_t)_{t \geq t_0}$  codant les *temps d'arrivée* du stimulus et pour sortie les données  $\mathbf{y}$ . Cette relation s'écrit :

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{Pl} + \mathbf{b}, \quad (1)$$

avec  $\mathbf{x}_{t_n} = [x_{t_n}, x_{t_n - \Delta t}, \dots, x_{t_n - P\Delta t}]^t \in \mathbb{R}^{P+1}$  et  $\mathbf{X} = [\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_N}]^t \in \mathbb{R}^{(P+1) \times N}$ . Dans le cas de protocoles *syn-*

chrones, les stimuli sont présentés aux instants d'acquisition  $t_n$  ( $\Delta t = TR$ ), si bien que  $\mathbf{X}$  est Toeplitz et donnée par sa première colonne. Elle perd cette caractéristique pour des protocoles *asynchrones* où le sujet est soumis à des stimuli intervenant aléatoirement ( $t'_n = n\Delta t$  avec  $\Delta t < TR$ ). Le pas  $\Delta t$  d'échantillonnage des colonnes de  $\mathbf{X}$  devient alors plus petit que  $TR$ , celui des lignes.

Dans (1), la matrice  $\mathbf{P} = [P_1, \dots, P_Q]$  est une base orthonormée de fonctions  $P_q = [P_q(t_1), \dots, P_q(t_N)]^t$  (e.g., base de sinus), fixée *a priori*, modélisant la dérive basse fréquence présente dans les données IRMf, traduisant un artefact lié au repliement spectral de phénomènes physiologiques (battements cardiaques, respiration).  $Q$  dépend de la fréquence minimale attribuable aux termes de dérive et  $l$  définit les coefficients de régression sur  $\mathbf{P}$  à estimer conjointement avec  $\mathbf{h}$ . Enfin,  $\mathbf{b}$  modélise le bruit d'acquisition, supposé blanc gaussien de variance inconnue  $\sigma_b^2 > 0$  et indépendant de  $\mathbf{h}$ .

## 2.2 Extensions

Les protocoles actuels d'IRMf enregistrent la réponse à une suite de stimuli (ou conditions) de différents types. Le modèle (1) s'adapte à cette situation en associant une FRH  $\mathbf{h}_m$  à chaque type de condition, modélisée par la séquence d'entrée  $(\mathbf{x}_i^{(m)})_{t \geq t_0}$  ou par  $\mathbf{X}^{(m)}$ . En effet, sous l'hypothèse d'un modèle de réponse *additif* vis-à-vis des filtres  $\mathbf{h}_m$ , l'équation (1) reste valide pour les paradigmes *multi-conditions* moyennant les adaptations suivantes :

$$\mathbf{X} = [\mathbf{X}^{(1)} | \dots | \mathbf{X}^{(M)}], \quad \mathbf{h} = [\mathbf{h}_1^t, \dots, \mathbf{h}_M^t]^t.$$

L'estimation du couple  $(\mathbf{h}, l)$  devient plus difficile car la taille de  $\mathbf{h}$  augmente tandis que celle de  $\mathbf{y}$  est inchangée. Toutefois, une expérience d'IRMf comporte plusieurs répétitions (ou *sessions*) du même paradigme. Le modèle (1) est mono-session ; dans le cas multi-session, il est naturel de considérer que  $\mathbf{h}$  reste *invariante* dès lors que le voxel, la condition et le sujet sont fixes. En notant  $i$  l'indice des sessions, le modèle (1) devient donc

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{h} + \mathbf{P}_i l_i + \mathbf{b}_i, \quad i \in \mathbb{N}_I^* \triangleq \{1, \dots, I\}, \quad (2)$$

qui tient compte des fluctuations des paramètres physiologiques ( $\mathbf{P}_i \in \mathbb{R}^{N_i \times Q_i}, l_i$ ) au cours des sessions, ainsi que du caractère pseudo-aléatoire de l'ordre des stimuli ( $\mathbf{X}_i$ ).

## 2.3 Vraisemblance des données

En supposant les réalisations du bruit  $(\mathbf{b}_i)_{1 \leq i \leq I}$  indépendantes et identiquement distribuées, la vraisemblance des données régies par le modèle (2) s'écrit

$$\begin{aligned} p(\mathbf{y} | \mathbf{h}, \mathbb{1}; \sigma_b^2) &= \prod_{i=1}^I (2\pi\sigma_b^2)^{-N_i/2} \exp\left(-\|\mathbf{y}_i - \mathbf{X}_i \mathbf{h} - \mathbf{P}_i l_i\|^2 / 2\sigma_b^2\right) \\ &= (2\pi\sigma_b^2)^{-N/2} \exp\left(-\|\mathbf{y} - \mathbb{X}\mathbf{h} - \mathbb{P}\mathbb{1}\|^2 / 2\sigma_b^2\right) \end{aligned}$$

où  $N = \sum_{i=1}^I N_i$ ,  $\mathbf{y} = [\mathbf{y}_1^t, \dots, \mathbf{y}_I^t]^t$ ,  $\mathbb{1} = [l_1^t | \dots | l_I^t]^t$ ,  $\mathbb{X} = [\mathbf{X}_1^t | \dots | \mathbf{X}_I^t]^t$  et  $\mathbb{P} = \text{diag}[\mathbf{P}_1, \dots, \mathbf{P}_I]$ .

Dans un cas pratique réaliste (voir Section 5), le nombre de paramètres est grand par rapport au nombre de données, si bien que l'estimation de  $(\mathbf{h}, \mathbb{1})$  au sens des moindres carrés s'avère peu robuste.

# 3 Estimateurs bayésiens de la FRH

## 3.1 Approche préexistante

Afin d'obtenir une estimée robuste de  $\mathbf{h}$ , la méthode proposée dans [4, 1] introduit des connaissances *a priori* sur la structure temporelle de chaque  $\mathbf{h}_m$ , sous la forme de contraintes aux bords  $\mathbf{h}_{m,0} = \mathbf{h}_{m,P\Delta t} = 0$  et d'un modèle *a priori* gaussien  $\mathcal{N}(\mathbf{0}, \sigma_{h_m}^2 \mathbf{R})$  où  $\mathbf{R} = (\mathbf{D}_2^t \mathbf{D}_2)^{-1}$  et  $\mathbf{D}_2$  est la matrice des différences d'ordre deux. L'indépendance *a priori* entre les conditions implique

$$\begin{aligned} p(\mathbf{h}; \boldsymbol{\theta}_H) &= \prod_{m=1}^M p(\mathbf{h}_m; \sigma_{h_m}^2) \\ &= |\mathbf{R}|^{-M/2} \prod_m (2\pi\sigma_{h_m}^2)^{(1-P)/2} \exp\left(-\mathbf{h}^t \mathbf{R}_H^{-1} \mathbf{h} / 2\right), \end{aligned}$$

où  $\mathbf{R}_H = \text{diag}[\sigma_{h_1}^2 \mathbf{R}, \dots, \sigma_{h_M}^2 \mathbf{R}]$  et  $\boldsymbol{\theta}_H = [\sigma_{h_1}^2, \dots, \sigma_{h_M}^2]$ .

L'estimateur de  $\mathbf{h}$  retenu dans [1] est celui du maximum *a posteriori* (MAP) calculé en maximisant la densité *a posteriori*  $p(\mathbf{h} | \mathbf{y}; \mathbb{1}, \boldsymbol{\theta}) \propto p(\mathbf{y} | \mathbf{h}; \mathbb{1}, \sigma_b^2) p(\mathbf{h}, \boldsymbol{\theta}_H)$ , soit

$$\hat{\mathbf{h}}^{\text{MAP}} = (\mathbb{X}^t \mathbb{X} + \sigma_b^2 \mathbf{R}_H^{-1})^{-1} \mathbb{X}^t (\mathbf{y} - \mathbb{P}\mathbb{1}), \quad (3)$$

qui dépend de  $\boldsymbol{\theta} = [\sigma_b^2, \boldsymbol{\theta}_H]$  et de  $\mathbb{1}$ , considérés comme inconnus mais fixes. La démarche adoptée dans [1] pour estimer  $(\boldsymbol{\theta}, \mathbb{1})$  consiste à minimiser l'anti-log-vraisemblance :

$$\mathcal{L}(\boldsymbol{\theta}, \mathbb{1}; \mathbf{y}) = -\ln p(\mathbf{y}; \mathbb{1}, \boldsymbol{\theta}) = \ln |\boldsymbol{\Upsilon}| + (\mathbf{y} - \mathbb{P}\mathbb{1})^t \boldsymbol{\Upsilon}^{-1} (\mathbf{y} - \mathbb{P}\mathbb{1})$$

avec  $\boldsymbol{\Upsilon} = r_b \mathbf{I}_N + \mathbb{X} \mathbf{R}_H \mathbb{X}^t$ , en recourant à une variante de l'algorithme EM [5]. Une fois la convergence locale atteinte,  $(\hat{\boldsymbol{\theta}}^{\text{MV}}, (\hat{\sigma}_b^2)^{\text{MV}}, \hat{\boldsymbol{\theta}}_H^{\text{MV}})$  remplace  $(\mathbb{1}, \sigma_b^2, \boldsymbol{\theta}_H)$  dans (3).

En choisissant  $\mathbb{1} = \hat{\mathbb{1}}^{\text{MV}}$ , cette approche ignore les fluctuations potentielles de  $\hat{\mathbb{1}}^{\text{MV}}$  qui peuvent être à l'origine d'un biais sur  $\hat{\mathbf{h}}^{\text{MAP}}$  et surtout sur les barres d'erreur calculées à partir de la covariance *a posteriori*. Nous proposons ici une alternative qui intègre cette variabilité.

## 3.2 Alternative

Nous modélisons  $\mathbb{1}$  comme un processus aléatoire indépendant de  $\mathbf{h}$  et régi par une distribution de probabilité  $p(\mathbb{1}; \boldsymbol{\theta}_{\mathbb{1}})$ , satisfaisant  $p(\mathbb{1}; \boldsymbol{\theta}_{\mathbb{1}}) = \prod_i p(l_i; \theta_{l_i})$ , où  $l_i \sim \mathcal{N}(\mathbf{0}, \sigma_{l_i}^2 \mathbf{I}_{Q_i})$ . Une version simplifiée consiste à poser  $\sigma_{l_i}^2 = \sigma_l^2, \forall i \in \mathbb{N}_I^*$ . En l'absence d'information *a priori* spécifique sur les dérivées, il est logique de choisir  $\sigma_l^2$  grand. Dans la suite, les expressions (4)-(7) définissant l'estimateur ainsi que les équations de l'EM (9)-(10) sont données après passage à la limite ( $\sigma_l^2 \rightarrow +\infty$ ).

En appliquant la règle de Bayes, on obtient la distribution *a posteriori* jointe  $p(\mathbf{h}, \mathbb{1} | \mathbf{y}; \boldsymbol{\theta}) \sim \mathcal{N}(\hat{\boldsymbol{\beta}}^{\text{MAP}}, \boldsymbol{\Sigma})$ , où  $\hat{\boldsymbol{\beta}}^{\text{MAP}} = [\hat{\mathbf{h}}^{\text{MAP}}, \hat{\mathbb{1}}^{\text{MAP}}]^t$  et  $\boldsymbol{\theta} = [\sigma_b^2, \boldsymbol{\theta}_H]$ . Le calcul de la matrice de covariance *a posteriori*

$$\boldsymbol{\Sigma} \triangleq \begin{bmatrix} \mathbf{R}_{H|\mathbb{Y}} & \mathbf{R}_{H\mathbb{1}|\mathbb{Y}} \\ \mathbf{R}_{H\mathbb{1}|\mathbb{Y}}^t & \mathbf{R}_{\mathbb{1}|\mathbb{Y}} \end{bmatrix}$$

requiert l'utilisation du lemme d'inversion de matrice partitionnée puisque  $\boldsymbol{\Sigma}^{-1}$  se lit à partir de  $p(\mathbf{h}, \mathbb{1} | \mathbf{y}, \boldsymbol{\theta})$ . Une application directe conduit à :

$$\mathbf{R}_{H|\mathbb{Y}} = \sigma_b^2 (\mathbb{X}^t (\mathbf{I}_N - \mathbb{P}\mathbb{P}^t) \mathbb{X} + \sigma_b^2 \mathbf{R}_H^{-1})^{-1}, \quad (4)$$

$$\mathbf{R}_{H\mathbb{1}|\mathbb{Y}} = -\mathbf{R}_{H|\mathbb{Y}} \mathbb{X}^t \mathbb{P}. \quad (5)$$

Pour la matrice  $\mathbf{R}_{\mathbb{L}|\mathbb{Y}}$ , il faut distinguer les termes diagonaux  $\mathbf{R}_{L_i|\mathbb{Y}} = \sigma_b^2 \mathbf{I}_{Q_i} + \mathbf{P}_i^t \mathbf{X}_i \mathbf{R}_{\mathbf{H}|\mathbb{Y}} \mathbf{X}_i^t \mathbf{P}_i$  des termes croisés  $\mathbf{R}_{L_i L_j|\mathbb{Y}} = \mathbf{P}_i^t \mathbf{X}_i \mathbf{R}_{\mathbf{H}|\mathbb{Y}} \mathbf{X}_j^t \mathbf{P}_j$ .

L'estimateur du MAP  $\hat{\boldsymbol{\beta}}^{\text{MAP}}$  s'obtient en maximisant  $p(\mathbf{h}, \mathbb{1} | \mathbb{y}; \boldsymbol{\theta})$  conjointement par rapport à  $\mathbf{h}$  et  $\mathbb{1}$  :

$$\hat{\mathbf{h}}^{\text{MAP}} = \mathbf{R}_{\mathbf{H}|\mathbb{Y}} \mathbb{X}^t (\mathbf{I}_N - \mathbb{P} \mathbb{P}^t) \mathbb{y} / \sigma_b^2, \quad (6)$$

$$\hat{\mathbb{1}}^{\text{MAP}} = \mathbb{P}^t (\mathbb{y} - \mathbb{X} \hat{\mathbf{h}}^{\text{MAP}}). \quad (7)$$

Les paramètres  $\boldsymbol{\theta}$  sont toujours estimés au sens du maximum de vraisemblance mais contrairement à [1], l'anti-log-vraisemblance ne dépend plus des dérivées  $\mathbb{1}$  :

$$\mathcal{L}(\boldsymbol{\theta}; \mathbb{y}) = -\ln p(\mathbb{y}; \boldsymbol{\theta}) = \ln |\boldsymbol{\Xi}| + \mathbb{y}^t \boldsymbol{\Xi}^{-1} \mathbb{y}$$

avec  $\boldsymbol{\Xi} = (\mathbf{I}_N / \sigma_b^2 - \mathbb{S} \boldsymbol{\Sigma} \mathbb{S}^t)^{-1}$  et  $\mathbb{S} = [\mathbb{X} | \mathbb{P}]$ . La minimisation de  $\mathcal{L}$  est itérative et coûteuse en temps de calcul à cause de  $\ln |\boldsymbol{\Xi}|$ . De plus,  $\mathcal{L}$  n'étant pas convexe, des minima locaux peuvent exister. Nous procédons néanmoins par optimisation locale à l'aide d'un algorithme EM.

## 4 Estimation des hyperparamètres

### 4.1 Principe de l'EM

L'algorithme EM, introduit par [6], est une méthode à point fixe qui garantit la minimisation locale de  $\mathcal{L}(\boldsymbol{\theta}; \mathbb{y})$  par rapport à  $\boldsymbol{\theta}$ . A partir d'une initialisation  $\boldsymbol{\theta}^0$ , une série d'estimées  $\boldsymbol{\theta}^k$  est générée en alternant les deux étapes suivantes :

$$\text{Expectation (E)} : \text{évaluer } \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^k), \quad (8a)$$

$$\text{Maximization (M)} : \boldsymbol{\theta}^{k+1} = \arg \max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^k), \quad (8b)$$

où  $\mathcal{Q}$  est définie comme l'espérance de la log-vraisemblance des « données complètes »,  $\mathbb{y}$  étant ici complété par  $(\mathbf{h}, \mathbb{1})$  :

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^k) &= \mathbb{E}_{\mathbf{H}, \mathbb{L}} [\ln p(\mathbb{y}, \mathbf{h}, \mathbb{1}; \boldsymbol{\theta}) | \mathbb{y}; \boldsymbol{\theta}^k] \\ &= \int \ln p(\mathbb{y}, \mathbf{h}, \mathbb{1}; \boldsymbol{\theta}) p(\mathbf{h}, \mathbb{1} | \mathbb{y}; \boldsymbol{\theta}^k) d\mathbf{h} d\mathbb{1}. \end{aligned}$$

### 4.2 Détails de l'algorithme

Puisque  $p(\mathbb{y}, \mathbf{h}, \mathbb{1}; \boldsymbol{\theta}) = p(\mathbb{y} | \mathbf{h}, \mathbb{1}; \sigma_b^2) p(\mathbf{h}; \boldsymbol{\theta}_{\mathbf{H}})$ , la fonction  $\mathcal{Q}$  admet la décomposition suivante :

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^k) = \mathcal{Q}_{\mathbb{Y}}(\sigma_b^2, \boldsymbol{\theta}^k) + \mathcal{Q}_{\mathbf{H}}(\boldsymbol{\theta}_{\mathbf{H}}, \boldsymbol{\theta}^k).$$

#### 4.2.1 Estimation de la variance du bruit $\sigma_b^2$

On montre assez directement que

$$\mathcal{Q}_{\mathbb{Y}} = -\frac{N}{2} \ln \sigma_b^2 - \frac{1}{2\sigma_b^2} \mathbb{E}[\|\mathbb{y} - \mathbb{X} \mathbf{h} - \mathbb{P} \mathbb{1}\|^2 | \mathbb{y}; \boldsymbol{\theta}^k].$$

Puisque  $p(\mathbf{h}, \mathbb{1} | \mathbb{y}; \boldsymbol{\theta}^k) \sim \mathcal{N}(\hat{\boldsymbol{\beta}}^{\text{MAP}}, \boldsymbol{\Sigma})$ , l'étape E (8a) s'écrit :

$$\mathbb{E}[\|\mathbb{y} - \mathbb{X} \mathbf{h} - \mathbb{P} \mathbb{1}\|^2 | \mathbb{y}; \boldsymbol{\theta}^k] = \|\mathbb{y} - \mathbb{S} \hat{\boldsymbol{\beta}}^{\text{MAP}}\|^2 + \text{tr}(\mathbb{S} \boldsymbol{\Sigma} \mathbb{S}^t).$$

Finalement, l'étape M (8b) pour  $\mathcal{Q}_{\mathbb{Y}}$  équivaut à identifier la valeur de  $\sigma_b^2$  pour laquelle  $\partial \mathcal{Q}_{\mathbb{Y}} / \partial \sigma_b^2 = 0$  :

$$\hat{\sigma}_b^2 = \frac{1}{N} (\|\mathbb{y} - \mathbb{S} \hat{\boldsymbol{\beta}}^{\text{MAP}}\|^2 + \text{tr}(\mathbb{S} \boldsymbol{\Sigma} \mathbb{S}^t)). \quad (9)$$

#### 4.2.2 Estimation des variances a priori $\sigma_{h_m}^2$

De même, l'étape E (8a) pour  $\mathcal{Q}_{\mathbf{H}}$  s'écrit :

$$\mathcal{Q}_{\mathbf{H}} = \frac{1-P}{2} \sum_m \ln \sigma_{h_m}^2 - \frac{M}{2} \ln |\mathbf{R}| - \mathbb{E}[\mathbf{h}^t \mathbf{R}_{\mathbf{H}}^{-1} \mathbf{h} | \mathbb{y}; \boldsymbol{\theta}^k],$$

avec

$$\mathbb{E}[\mathbf{h}^t \mathbf{R}_{\mathbf{H}}^{-1} \mathbf{h} | \mathbb{y}; \boldsymbol{\theta}^k] = (\hat{\mathbf{h}}^{\text{MAP}})^t \mathbf{R}_{\mathbf{H}}^{-1} \hat{\mathbf{h}}^{\text{MAP}} + \text{tr}(\mathbf{R}_{\mathbf{H}|\mathbb{Y}} \mathbf{R}_{\mathbf{H}}^{-1}).$$

L'étape M (8b) pour  $\mathcal{Q}_{\mathbf{H}}$  nous conduit donc à :

$$\hat{\sigma}_{h_m}^2 = \frac{1}{P-1} \text{tr} \left\{ \left( \hat{\mathbf{h}}_m^{\text{MAP}} (\hat{\mathbf{h}}_m^{\text{MAP}})^t + \mathbf{R}_{\mathbf{H}_m|\mathbb{Y}} \right) \mathbf{R}_{\mathbf{H}}^{-1} \right\}. \quad (10)$$

L'algorithme consiste donc à itérer jusqu'à convergence les remises à jour de  $\boldsymbol{\Sigma}$  (cf. (4)-(5)), le calcul de  $\hat{\boldsymbol{\beta}}^{\text{MAP}}$  et celui de  $\boldsymbol{\theta}$  à l'aide respectivement de (6)-(7) et (9)-(10).

## 5 Évaluation des performances

Nous comparons les performances des estimateurs  $\hat{\mathbf{h}}^{\text{MAP}}$ , définis en (3) et (6), en termes de biais et de variance d'estimation (barres d'erreur).

### 5.1 Simulation

Pour chaque session, nous simulons une séquence *aléatoire* mélangeant les indices de  $M = 2$  types de condition. Les intervalles inter-stimuli sont aléatoires et suivent une distribution uniforme sur  $[2,5; 3,5]$  s. Les temps d'arrivée des stimuli sont placés sur une grille espacée de  $\delta t = 0,5$  s. Les séquences binaires  $\mathbf{x}_i^m$  couplées aux conditions  $m = 1; 2$  sont ensuite convoluées par une FRH,  $\mathbf{h}_1$  ou  $\mathbf{h}_2$ , reportées en Fig. 1a. Il apparaît que  $\mathbf{h}_2$  est plus piquée que  $\mathbf{h}_1$  et a une largeur à mi-hauteur plus faible. Un bruit blanc gaussien ( $\sigma_b^2 = 0,45$ ; RSB = 17 dB) a été ajoutée, ainsi qu'une dérive basse fréquence  $\mathbf{P}_i \mathbf{l}_i$ , aléatoire et de période de coupure variable entre 80 et 120 secondes selon les  $I = 4$  sessions ( $Q_i \in \{6, \dots, 9\}$ ). Les données sont enfin obtenues par sous-échantillonnage de  $\mathbf{x}_i^m * \mathbf{h}_m$  au taux  $F = [TR/\delta t]$  ( $TR = 2$  s), et sommation sur  $m$ .

Les Fig. 1b-c montrent que l'estimateur (6) fournit des solutions plus proches de  $\mathbf{h}_1$  et  $\mathbf{h}_2$  de 30% environ par rapport à (3) en termes de distance quadratique, dont les barres d'erreur, directement proportionnelles à la diagonale de  $\boldsymbol{\Sigma}$ , majorent celles calculées pour (3) de 5 à 10 % selon les échantillons temporels. Ceci traduit une variance d'estimation légèrement accrue, donc une certaine prise en compte de la variabilité de  $\hat{\mathbb{1}}^{\text{MV}}$ . Les propriétés statistiques de  $\hat{\mathbf{h}}^{\text{MAP}}$  ont été étudiées quantitativement par une étude de Monte-Carlo portant sur 1000 réalisations de bruit. Cette étude démontre une réduction du biais et de l'erreur en moyenne quadratique de 5 % et 15 % environ pour  $\mathbf{h}_1$  et  $\mathbf{h}_2$ , respectivement.

En terme de coût de calcul, l'algorithme EM proposé est comparable à l'ECM de [1] : 25 itérations en 1,5 s, contre 130 itérations en 0,95 s sur un Pentium IV 1,8 GHz.

### 5.2 Données réelles

Cette nouvelle approche a été également testée sur des données issues d'une expérience sur la compréhension du langage.

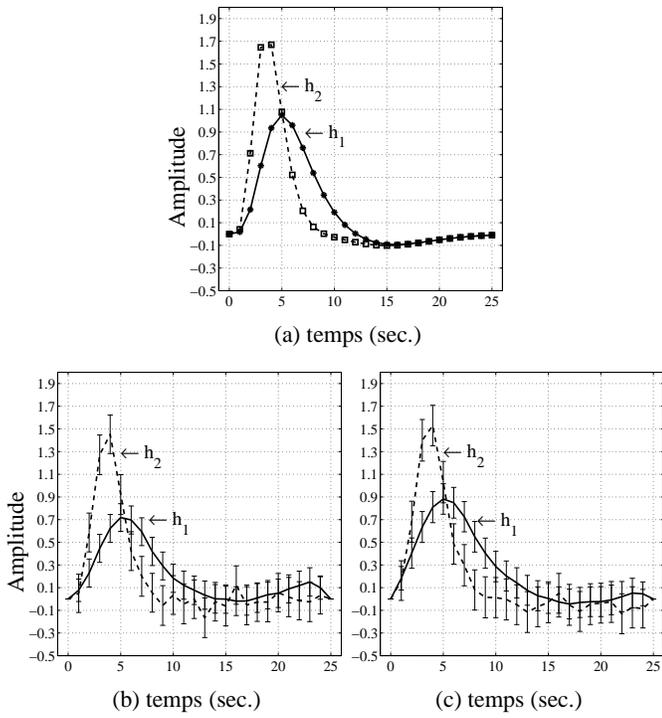


FIG. 1 – (a) : Vraies FRHs. (b) : FRHs calculées à partir de (3). (c) : FRHs correspondant à (6).

Dans cette expérience, chaque stimulus est composé de deux pseudo-mots délivrés par un casque audio. La tâche du sujet consiste à indiquer s’il perçoit une différence entre eux. Trois conditions ( $M = 3$ ) sont présentes : une « phonologique » ( $h_1$ ), une « acoustique » ( $h_2$ ) et une de « contrôle » ( $h_3$ ). Dans la condition « phonologique », la différence entre les stimuli met en jeu un contraste utilisé pour distinguer des mots dans la langue du sujet. Dans la condition « acoustique », le contraste entre les stimuli n’est pas pertinent dans la langue du sujet, mais peut l’être dans une langue étrangère. Dans la condition « contrôle », les deux pseudo-mots sont identiques. Chaque essai dure 2 s auxquelles succède une acquisition du volume cérébral complet pendant 1,3 s à l’aide d’une séquence EPI ( $TR = 3,3$  s). L’expérience comporte six sessions ( $I = 6$ ) de  $N_i = 100$  scans chacune.

Les résultats de la Fig. 2 appellent deux commentaires. D’une part, les FRH estimées ont des formes similaires pour les trois conditions, mais la discrimination entre ( $h_1, h_2$ ) d’une part, et  $h_3$  d’autre part semble facilitée en recourant à l’approche de la Section 3.2. Dans les deux cas (Fig. 2a et Fig. 2b), ces résultats confirment l’hypothèse cognitive sous-jacente selon laquelle, le Gyrus Supra Marginal est une aire du langage non spécifique à la langue maternelle du sujet. En effet, comme le montre l’amplitude du pic de  $\hat{h}_1$  et  $\hat{h}_2$  sur les Fig. 2ab, le sujet perçoit dans cette région une différence entre les stimuli (ou pseudo-mots), que celle-ci soit soit phonologique ou acoustique.

D’autre part, les barres d’erreur de la Fig. 2b majorent de 15 % environ celles de la Fig. 2a, traduisant une prise en compte de la variabilité de l’estimation de  $\mathbb{1}$ , jusque-là négligée en faisant  $\mathbb{1} = \hat{\mathbb{1}}^{MV}$ .

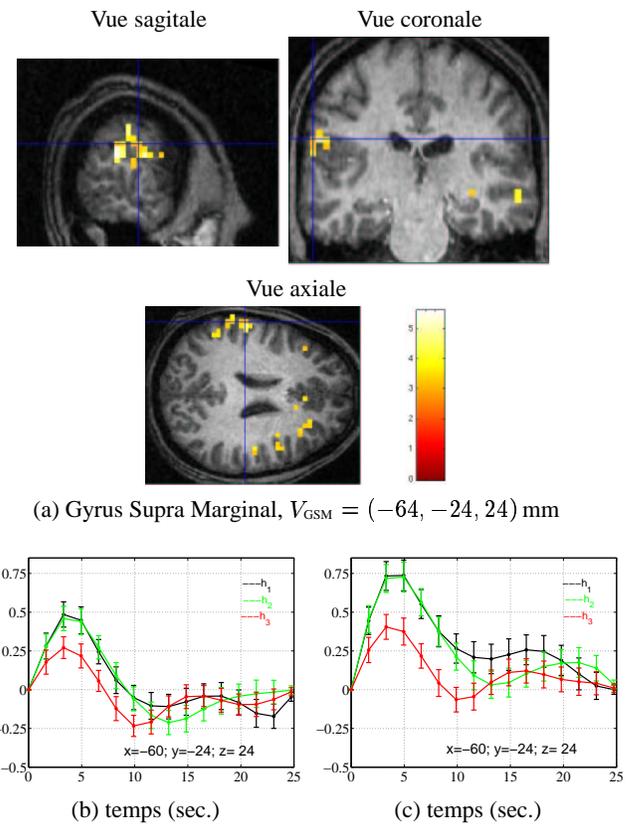


FIG. 2 – (a) Activations dans le gyrus supra marginal détectée à  $p = 10^{-3}$  pour le contraste « (phonologique-contrôle)&(acoustique-contrôle) ». (b) et (c) : ( $h_1, h_2, h_3$ ) calculées en  $V_{GSM}$  par (3) et (6), respectivement.

## Références

- [1] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, C. Pallier et H. Benali, « Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment », *IEEE Trans. Medical Imaging in press*, 2003.
- [2] S. Ogawa, T. Lee, A. Kay et D. Tank, « Brain magnetic resonance imaging with contrast dependent on blood oxygenation », *Proc. Natl. Acad. Sci. USA*, vol. 87, n° 24, pp. 9868–9872, 1990.
- [3] P. A. Bandettini, A. Jesmanowicz, E. C. Wong et J. S. Hyde, « Processing strategies for time-course data sets in functional MRI of the human brain », *Magn. Reson. Med.*, vol. 30, pp. 161–173, 1993.
- [4] G. Marrelec, H. Benali, P. Ciuciu, M. Péligrini-Issac et J.-B. Poline, « Robust Bayesian estimation of the hemodynamic response function in event-related BOLD MRI using basic physiological information », *Hum. Brain Mapp.*, vol. 19, n° 1, pp. 1–17, mai 2003.
- [5] X.-L. Meng et D. B. Rubin, « Maximum likelihood estimation via the ECM algorithm : A general framework », *Biometrika*, vol. 80, pp. 267–278, 1993.
- [6] A. P. Dempster, N. M. Laird et D. B. Rubin, « Maximum likelihood from incomplete data via the EM algorithm », *J. R. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.