

# Modèles Semi-Markoviens Cachés de Séquences d'Événements : Application en Cartographie IRM Fonctionnelle Cérébrale

Sylvain FAISAN<sup>1,2</sup>, Laurent THORAVAL<sup>1</sup>, Jean-Paul ARMSPACH<sup>2</sup>

<sup>1</sup> LSIIT/GRI - UMR CNRS 7005,  
ENSPS, Bd. Sébastien Brant  
F-67400 Illkirch

<sup>2</sup>Institut de Physique Biologique,  
Faculté de Médecine - UMR CNRS 7004,  
4, Rue Kirschleger, F-67085 Strasbourg

{faisan,thoraval}@ensps.u-strasbg.fr, armspach@ipb.u-strasbg.fr

**Résumé** – L'hypothèse de stationarité par morceaux du processus observable d'une chaîne de Markov cachée rend les modèles de Markov cachés (MMCs) inadaptés à l'analyse de processus aléatoires de nature événementielle. Pour de tels processus, une nouvelle classe de MMCs est proposée : les modèles semi-Markoviens cachés de séquences d'événements. Le processus observable n'est plus de nature segmentale mais issu d'une étape de détection et de caractérisation. Le formalisme markovien standard est adapté en conséquence. Les résultats obtenus en cartographie de séquence d'images IRM fonctionnelle valident cette nouvelle approche de modélisation statistique tout en ouvrant de nouvelles perspectives dans la reconnaissance et la détection de processus aléatoires de nature événementielle.

**Abstract** – Due to the piecewise stationarity assumption required for the observable process of a hidden Markov chain, the application of Hidden Markov Models (HMMs) to the analysis of event-based random processes remains intricate. For such processes, a new class of HMMs is proposed: the Hidden Semi-Markov Event Sequence Model (HSMESM). In a HSMESM, the observable process is no more considered as segmental in nature but issued from a detection-characterization preprocessing step. The standard markovian formalism is adapted accordingly. Results obtained in functional MRI sequence analysis validate this novel statistical modeling approach while opening new perspectives in detection-recognition of event-based random processes.

## 1 Introduction

De par l'hypothèse de stationarité par morceaux du processus observable qui les compose, les modèles de Markov cachés (MMCs) s'avèrent particulièrement adaptés à l'analyse de processus aléatoires de nature segmentale. Cependant, nombre de processus, rencontrés notamment en génie biomédical, ne sont pas de cette nature. Ils se présentent sous la forme d'une séquence d'événements localisés dans le temps, à caractère le plus souvent non stationnaire, et noyés dans du bruit ou toute autre activité de fond : ondes ECG (P, QRS, T) en électrocardiographie, pointes-ondes en électroencéphalographie, ruptures de dynamique en imagerie IRM fonctionnelle cérébrale. Pour ce type de processus, une modélisation statistique par chaîne de Markov cachée de la distribution temporelle des événements d'intérêt reste toujours pertinente. Par contre, l'élaboration comme l'obtention de segments d'observations à la fois stationnaires sur le support temporel de ces événements et discriminants entre classes d'événements se révèlent difficiles, voire impossibles, parce qu'incompatibles avec la nature événementielle de ces processus. De facto, les performances d'analyse de tels processus par MMCs standards s'en trouvent dégradées. Pour résoudre ce problème, nous proposons d'adapter le formalisme markovien caché en le plaçant d'emblée dans un cadre de détection et d'analyse de séquences d'événements. La redéfinition nécessaire du contenu de la séquence d'obser-

vations conduit à développer un nouveau type de MMC pour l'analyser : le modèle semi-markovien caché de séquence d'événements (MSMCSE). Conçu sur la base d'un modèle semi-markovien caché (MSMC), le MSMCSE se révèle mieux adapté à l'analyse de processus aléatoires de nature événementielle plutôt que segmentale. Après avoir défini dans une première partie les aspects formels de ces nouveaux modèles, nous présentons ensuite dans une seconde partie une application des MSMCSEs en cartographie IRM fonctionnelle cérébrale.

## 2 Le modèle semi-Markovien caché de séquences d'événements

### 2.1 Séquence d'événements

Une étape de prétraitement détecte et caractérise les événements d'intérêt composant le processus événementiel à analyser. A chaque instant de détection  $t_i$ , une observation, encore appelée événement,  $e_i$ , est produite. Soient  $\tau_O = \{1, 2, \dots, T\}$  l'ensemble des instants d'observation et  $O = o_1 o_2 \dots o_T$  la séquence d'observation correspondante. Soient  $\tau_E = \{t_1 = 1, t_2, \dots, t_L = T\}$ ,  $\tau_E \subset \tau_O$ , l'ensemble des instants de détection d'événements et  $E = e_1 e_2 \dots e_L$  la séquence des  $L$  événements détectés correspondante.  $e_1$  et  $e_L$  sont deux événements fictifs ajoutés artificiellement aux extrémités de la séquence  $E$

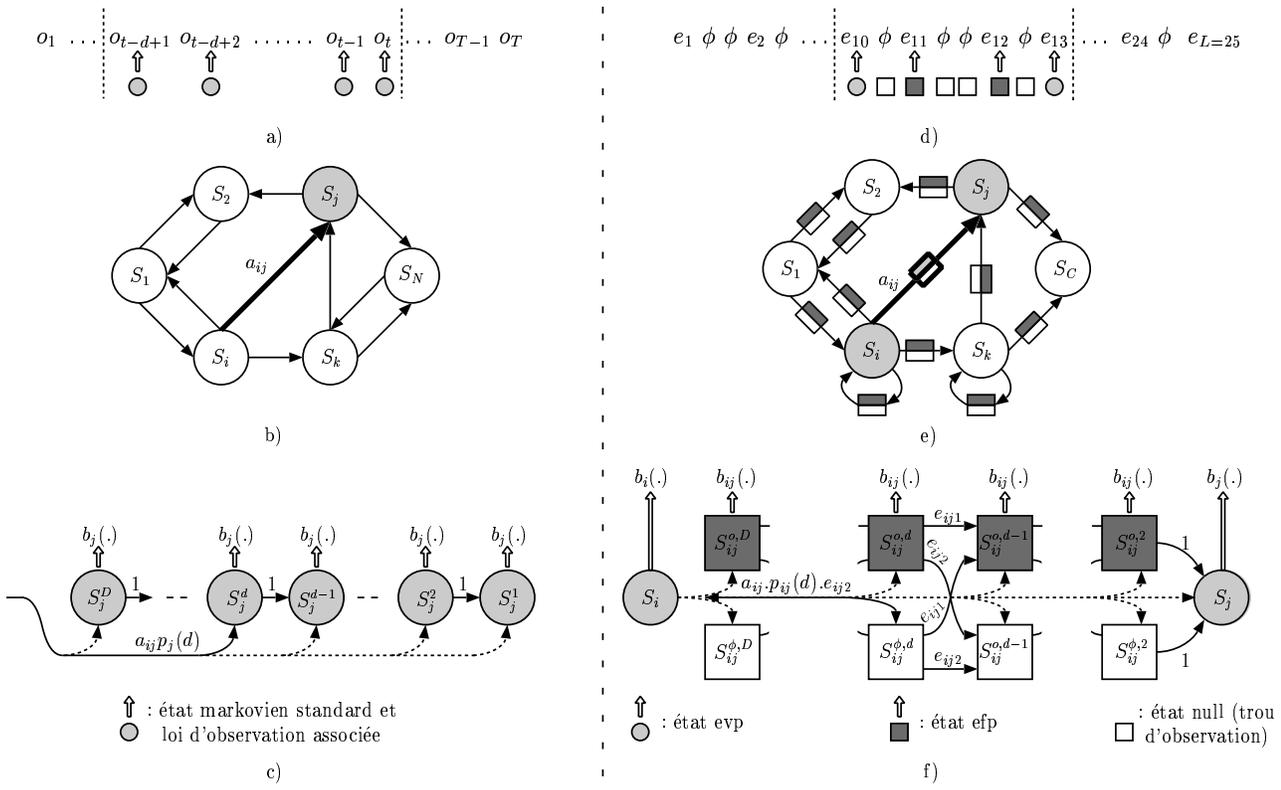


FIG. 1 – Structure des séquences d’observations d’un MSMC (a) et d’un MSMCSE (d). Représentation compacte d’un MSMC (b) et d’un MSMCSE (e). Représentation markovienne standard équivalente de l’état semi-markovien  $S_j$  venant de  $S_i$  (c) et de l’état semi-markovien  $S_{ij}$  intercalé entre les états markoviens  $S_i$  et  $S_j$  (f).

pour des raisons de modélisation des durées inter-événements d’intérêt. Alors, par définition, la séquence  $O$  est déduite de la séquence  $E$  par insertion d’un événement *null*,  $\phi$ , dans  $E$  pour chaque  $t \in \tau_{\bar{E}} = \tau_O - \tau_E$ . L’égalité  $o_t = \phi$  signifie qu’un trou d’observation, résultat d’une non détection, apparaît à l’instant  $t$ . A titre d’exemple, si  $T = 20$  et  $\tau_E = \{1, 3, 6, 12, 15, 20\}$ , alors  $E = e_1 e_2 e_3 e_4 e_5 e_6$ , conduisant à la séquence d’observations  $O = e_1 \phi e_2 \phi \phi e_3 \phi \phi \phi \phi e_4 \phi \phi e_5 \phi \phi \phi \phi e_6$ .

## 2.2 Eléments d’un MSMCSE

De par la nature orientée détection du prétraitement appliqué au processus à analyser, toute séquence d’observations d’un MSMCSE voit potentiellement constituée d’événements vrais positifs (*evp*), caractérisant le processus, ainsi que d’événements faux positifs (*efp*) et de trous d’observation (*null*). Si  $C - 2$  représente le nombre de classes d’*evp* caractérisant le processus modélisé, alors l’espace d’états du MSMCSE associé résulte de l’union des sous-espaces d’états :

- $S_{evp} = \{S_i; i = 1, \dots, C\}$  : les  $S_i$  sont des états markoviens cachés standards produisant des *evp* aux instants  $t_i \in \tau_E$  suivant leur loi d’observation propre  $b_i(o_{t_i})$ .  $S_1$  et  $S_C$  sont les états de départ et d’arrivée du processus caché ; ils modélisent les *evp* fictifs  $e_1$  et  $e_L$  avec arbitrairement  $b_1(o_{t_1} = e_1) = 1$  et  $b_C(o_{t_L} = e_L) = 1$ .
- $S_{efp/null} = \{S_{ij}; i, j = 1, \dots, C\}$  : les  $S_{ij}$  sont des états semi-markoviens cachés intercalés entre les  $S_i$ . Si  $d$  ( $d \geq 1$ ) est la durée entre deux états  $S_i$  et  $S_j$  successifs, lors du transit de  $S_i$  à  $S_j$ , l’état  $S_{ij}$  est occupé  $d - 1$  fois.

La durée  $d$  est modélisée explicitement par la loi  $p_{ij}(d)$  conformément au caractère semi-markovien de l’état  $S_{ij}$ . Les  $S_{ij}$  produisent des *efp* ou des *null* aux instants  $t \in \tau_{\bar{E}}$  suivant leur loi composite propre  $b_{ij}^+(o_t) = b_{ij}(e_l)$ .  $e_{ij1} \cdot \mathbb{1}_{\tau_E}(t) + e_{ij2} \cdot \mathbb{1}_{\tau_{\bar{E}}}(t)$  ( $\mathbb{1}_X(t) = 1$  si  $t \in X$ , 0 sinon), construite sur le processus de Bernoulli de paramètre  $e_{ij1} = 1 - e_{ij2}$ .

Un MSMCSE est entièrement défini par la donnée du paramètre  $C$  et du vecteur paramètre  $\lambda = (A, B, B', D, E)$  où  $A = \{a_{ij}\}$  désigne la matrice de probabilités de transition,  $B = \{b_j(\cdot)\}$  et  $B' = \{b_{ij}(\cdot)\}$  sont les lois d’observation,  $D = \{p_{ij}(\cdot)\}$  sont les lois de durée et  $E = \{e_{ij1}\}$  la matrice de probabilités d’émission, représentative des différents taux de fausses alarmes entre classes d’événements d’intérêt.

D’un point de vue synthèse, un MSMCSE  $\lambda$  génère la séquence d’observations  $O$  comme suit. En  $S_1$  à  $t_1$ , le processus caché émet  $e_1$ . A  $t_k$ ,  $1 \leq k < L$ , en  $S_i$ , le processus choisit le prochain état  $S_j$  à visiter suivant la distribution de probabilité de transition  $\{a_{ij}\}$ .  $S_j$  est effectivement visité à  $t_l = t_k + d$ ,  $l > k$ , avec  $d$  tiré aléatoirement suivant la loi de durée inter-*evp*  $p_{ij}(d)$ . Avant d’atteindre  $S_j$ , le processus reste  $d - 1$  fois dans l’état  $S_{ij}$  pour y émettre successivement des *efp* ou des *null* suivant la loi  $b_{ij}^+(\cdot)$ . Cette procédure est répétée jusqu’à atteindre l’état absorbant  $S_C$  à  $t_L = T$  pour y émettre  $e_L$ .

### 2.3 Représentation markovienne standard équivalente : du MSMC au MSMCSE

La représentation markovienne standard équivalente d'un MSMCSE, illustrée partie droite de la figure 1, peut être obtenue à partir de celle d'un MSMC établie par Levinson [4], représentée partie gauche de la même figure, en appliquant à cette dernière les transformations topologiques suivantes : *i*) renoter  $S_j^1$  en  $S_j$ , *ii*) dupliquer  $S_j^d$  en  $S_j^{o,d}$  et  $S_j^{\phi,d}$ ,  $D \leq d \leq 2$ , *iii*) réindicer  $S_j^{\cdot d}$  en  $S_j^{\cdot,d}$ ,  $D \leq d \leq 2$ , *iv*) associer  $b_{ij}(\cdot)$  à  $S_j^{o,d}$ ,  $D \leq d \leq 2$ , *v*) interconnecter les états  $S_i, S_j, S_j^{o,d}$  et  $S_j^{\phi,d}$  avec les probabilités de transition appropriées, comme indiqué figure 1.f. Chaque état semi-markovien  $S_{ij}$  peut donc être vu comme deux lignes à retard composées d'états markoviens en parallèle entre lesquelles le processus caché transite suivant les probabilités  $e_{ij1}$  et  $e_{ij2}$ . Observer le segment d'observations de la figure 1.d revient au niveau de la chaîne de Markov à occuper successivement  $S_i, S_j^{\phi,7}$ , puis  $S_j^{o,6}$  et ainsi de suite jusqu'à atteindre  $S_j$ .

### 2.4 Aspects algorithmiques

Les 3 problèmes fondamentaux que sont l'évaluation, l'apprentissage et le décodage [2] classiquement rencontrés dans le cadre des MMCs ou des MSMCs se posent et se résolvent de manière analogue dans le cas des MSMCSEs. Leur résolution conduit à des versions modifiées de l'algorithme Forward-Backward, de l'algorithme de Baum-Welch et de l'algorithme de Viterbi, respectivement. Nous ne les présentons pas ici. Soulignons toutefois que dans une approche par MSMCSEs, les variables Forward, Backward et de Viterbi, ne sont évaluées que pour la suite d'instant de détection  $t_l$ , réduisant sensiblement la charge de calcul ainsi que les ressources mémoires nécessaires dès lors que  $L \ll T$ .

## 3 Application en cartographie IRMf cérébrale

L'IRM fonctionnelle (IRMf) cérébrale permet d'analyser in vivo, dans l'espace et dans le temps, les phénomènes neuro-naux, métaboliques, et hémodynamiques liés à l'activité cérébrale. La cartographie IRMf cérébrale vise à délimiter les zones fonctionnelles cérébrales activées par une tâche motrice, sensorielle ou cognitive spécifique appliquée au patient conformément à un paradigme d'activation prédéfini. La détection d'activité s'appuie sur les résultats obtenus par traitement du signal IRMf associé à chaque voxel. En cas d'activité, le signal IRMf présente des variations d'amplitude en corrélation avec les transitions repos-activation-repos du paradigme. Les principales méthodes de cartographie hypothèquent sur la forme du signal IRMf actif attendu [5]. La méthode que nous proposons se limite à l'analyse de la séquence d'évènements d'intérêt que sont les ruptures de dynamique du signal IRMf, minimisant ainsi le nombre d'hypothèses a priori. Les ruptures du signal sont détectées au moyen d'une analyse multirésolution par ondelettes [1],[3]. Le taux d'activité cérébrale locale est évalué à partir du degré d'alignement temporel entre la séquence de ruptures détectées et le MSMCSE associé au paradigme d'activation.

### 3.1 Modélisation par MSMCSE du paradigme d'activation, Apprentissage

Le MSMCSE d'activation  $\lambda$  est construit à partir de la séquence déterministe de transitions repos-activation (OFF-ON) du paradigme de stimulation (Fig.2). A chaque transition OFF-

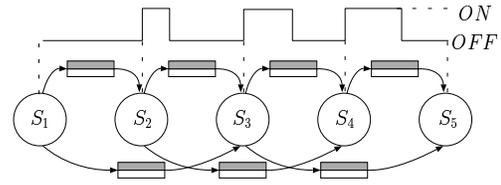


FIG. 2 – haut : Séquence de transitions du paradigme d'activation, bas : Chaîne de Markov cachée du MSMCSE correspondant.

ON est associé un état *evp*  $S_i$  d'indice  $i$  représentatif de son ordre d'apparition dans le paradigme. L'espace d'états du modèle est complété par les états initial et final puis par les états semi-markoviens  $S_{ij}$ . La topologie de la chaîne est de type gauche-droite, avec pour contrainte supplémentaire  $a_{ij} = 0$  si  $j > i + \Delta$ ,  $\Delta = \frac{C}{2}$ . Les lois d'observation et les lois de durée sont considérées gaussiennes. Le concept de liaison de paramètres [2] est utilisé pour réduire la taille du vecteur paramètre du modèle.

L'apprentissage du MSMCSE  $\lambda$  est effectué au moyen de la procédure itérative de la figure 3. A partir d'un modèle initial

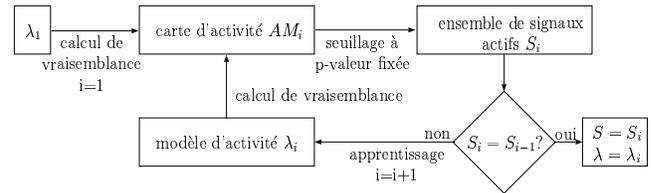


FIG. 3 – Procédure d'apprentissage du MSMCSE (voir texte).

$\lambda_1$ , une carte d'activité est calculée en déterminant pour chaque voxel la vraisemblance  $l = P(O|\lambda)$ . Une p-valeur  $p$  est ensuite associée à chaque vraisemblance  $l$ . Par définition,  $p = \int_l^{+\infty} P(x|H_0)dx$  où  $P(x|H_0)$  est la densité de probabilité de la vraisemblance  $x$  sous l'hypothèse  $H_0$  (inactivité cérébrale). Cette distribution est inférée expérimentalement en générant aléatoirement des séquences d'évènements représentatives de  $H_0$  à partir de l'ensemble des séquences d'évènements intracranienne. Après seuillage de la carte, les signaux déclarés actifs servent de nouvelle base d'apprentissage pour la réestimation des paramètres du modèle. La procédure est répétée jusqu'à convergence.

### 3.2 Résultats sur séquences IRMf

L'approche par MSMCSE a été appliquée sur des séquences d'images IRMf cérébrales 3-D pour un paradigme d'activation des zones fonctionnelles du langage. Les images, acquises avec un imageur S200 Bruker 2 Teslas (antenne de volume, EPI, orientation axiale, 32 coupes,  $64 \times 64$  pixels, volume voxel =  $4 \times 4 \times 4$  mm, TE=10ms, TR=5s), ont été recalées sur

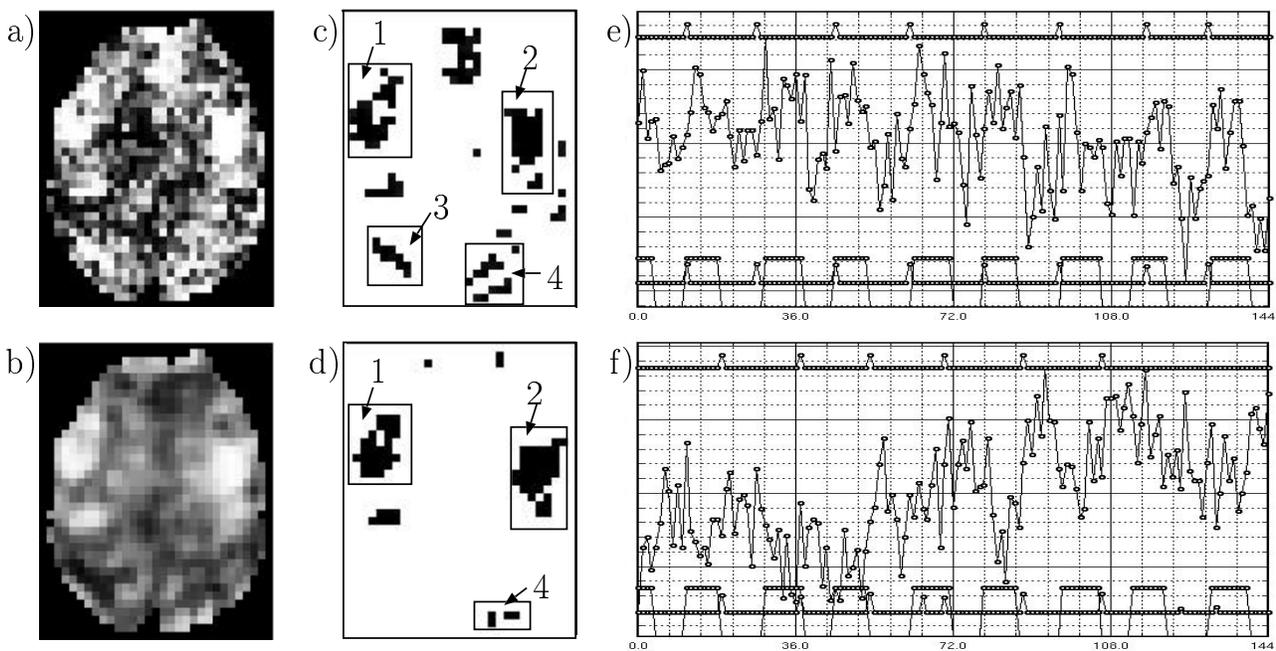


FIG. 4 – a,b) Cartes d’activités en terme de p-valeur (p-valeur=0 (blanc), p-valeur=1 (noir)) obtenues par la méthode MSMCSE (a) et le  $t$ -test (b). c,d) Cartes d’activités seuillées correspondantes. e,f) Exemples de signaux actifs détectés par les deux méthodes (e) et uniquement par la méthode MSMCSE (f), avec, pour chaque figure, de bas en haut : paradigme d’activation, évènements détectés, signal IRMF, instants d’occurrence des *evp* obtenus par décodage de Viterbi.

la première image de la série avant analyse. Les cartes d’activité obtenues par l’approche MSMCSE ont été comparées avec celles du  $t$ -test, méthode utilisée communément en routine clinique. Un échantillon de ces résultats est donné figure 4. Les résultats montrent la capacité de la méthode développée à détecter des signaux actifs non détectés par le  $t$ -test. Les zones 1,2,3 et 4 de la figure 4.c correspondent à des régions fonctionnelles impliquées dans le langage et validées par un expert. On remarque que la zone 3 n’a pas été détectée par le  $t$ -test. Il convient d’ajouter que les signaux détectés actifs hors des zones du langage ne doivent pas être considérés pour autant comme des fausses alarmes, le paradigme d’activation pouvant induire des activités cérébrales secondaires. De façon générale, les signaux détectés par la méthode MSMCSE mais non détectés par le  $t$ -test présentent soit une dérive de ligne de base, soit un bruit impulsionnel important, soit un retard d’activité (cas de la région 3 et de celle située entre les régions 1 et 2).

## 4 Conclusion

Les processus aléatoires événementiels pour lesquels l’information d’intérêt est portée par des évènements non-stationnaires localisés temporellement restent difficilement modélisables par une approche markovienne standard. L’approche par MSMCSE propose d’adapter le processus observable au caractère événementiel et non stationnaire de ces processus tout en préservant une modélisation statistique efficace par chaîne de Markov de leur distribution temporelle.

Sur le plan applicatif, les résultats obtenus en cartographie IRMF cérébrale permet de valider l’approche, novatrice au re-

gard du nombre limité d’hypothèses avancées quant à la morphologie du signal IRMF actif attendu.

## Références

- [1] Aldroubi A. et Unser M. : Wavelets in Medicine and Biology. CRC Press, 1996.
- [2] Rabiner L.R. : A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, vol. 77, no. 2, pp 257–286, Février 1989.
- [3] Thoraval L., Armspach J.-P. et Nammer I. : Analysis of brain functional MRI time series based on continuous wavelet transform and stimulation-response coupling distance. In Proc. MICCAI Conf, pp. 881–888, Octobre 2001.
- [4] Levinson S.E. : Continuously variable duration hidden Markov models for automatic speech recognition. Computer Speech and Language, vol. 1, no.1, pp.29–45, Mars 1986.
- [5] Friston K.J., Jezzard P. et Turner R. : Analysis of functional MRI time-series. Human Brain Mapping, vol. 1, pp.153–171, 1994.