HIDDEN MARKOVIAN MODELING AND ANALYSIS OF MULTIPLE-EVENT-SEQUENCE-BASED RANDOM PROCESSES. APPLICATION TO ROBUST DETECTION OF BRAIN FUNCTIONAL ACTIVATION

S. Faisan, L. Thoraval, F. Heitz

LSIIT, UMR CNRS-ULP 7005,
Strasbourg I University,
67412 Illkirch, France

J.-P. Armspach

IPB, UMR CNRS-ULP 7004,
Strasbourg I University,
67085 Strasbourg, France

ABSTRACT

This paper presents a novel statistical approach for the modeling and analysis of structured random processes observed through multiple event sequences: the hidden Markov multiple event sequence model (HMMESM). This model accounts for several features of these processes: (i) the hidden–observable aspect of the event sequences to be analyzed, (ii) the multiplicity of the observed event sequences, (iii) the non stationary, time-localized character of their events, (iv) the redundancy, complementarity, and strong asynchrony that exist between events across sequences. A first application of this model in functional MRI (fMRI) brain mapping is presented. The developed method shows high robustness to noise and variability of the active fMRI signals.

1. INTRODUCTION

Many random processes, met in particular in the biomedical field, appear as a hidden structured event-based process observed through multiple correlated sequences of highly nonstationary, time-located events. For example, the cardiac electrophysiological process can be observed through multiple sequences of electrocardiographic events (P, QRS, T), whereas the neuronal activation process in functional MRI (fMRI) can be observed through multiple sequences of dynamics changes detected in the fMRI signals. Unfortunately, three main difficulties arise when analyzing such processes, when the standard hidden Markov modeling (HMM) framework is used. The piecewise stationarity assumption required for the observable process of an HMM is incompatible with the transient, nonstationary nature of the observed events. The HMM formalism is not well suited to handle simultaneously a large number of event sequences and to take into account strong asynchrony that may exist between events across sequences. To cope with these limitations, we propose to adapt the HMM formalism by placing it within the scope of event detection and event sequence fusion. In this context, a preprocessing step detects, independently in each observable channel, events of interest. Detected events are then associated across channels, based on knowledge about the hidden process under analysis and on causality constraints between event associations. The new meaning thus given to the sequence of observations leads to a new type of HMMs : the hidden Markov multiple event sequence model (HMMESM). By exploiting the redundancy and the complementarity of events detected across multiple observation channels, this model is well adapted to the analysis of multiple event-based random processes. As an illustration, activation detection results obtained by HMMESMs in fMRI brain mapping are particularly convincing.

2. HIDDEN MARKOV MULTIPLE EVENT SEQUENCE MODELS

An HMMESM is a doubly stochastic process, $\{X_u, \boldsymbol{O}_u\}$, dedicated to the modeling and analysis of a non directly observable, event-based random process, the so-called *deep* process, observed through N correlated event-based random processes, the *shallow* processes. The term “correlated” means here that a deep event gives rise to N observable events at most, one per process, in the shallow processes. Also, to be observable, and thereby detectable, any deep event should give rise to one shallow event at least. In this context, $\{X_u\}$, the hidden part of an HMMESM, models the ordering of deep events along the ordering axis u. Its observable counterpart, $\{\boldsymbol{O}_u\}$, accounts for the temporal aspects and the short-term statistical characteristics of the deep and related shallow events along the time axis t.

2.1. Multiple event sequences, Scenarios, Observation sequence

In an HMMESM approach, a preprocessing step detects, independently in each input shallow process, events of interest for the deep process analysis. Let $I = [0, T]$ be the observation time interval, N be the number of shallow processes, $\boldsymbol{\rho}^{(k)}$, $1 \leq k \leq N$, be the occurrence time set of the events detected in the k-th shallow process. Two fictive events are introduced at the beginning and the end of each detected event sequence for modeling purposes. For the sake of simplicity, an event is only characterized by its occurrence time so that the detected event sequences are finally denoted: $\boldsymbol{E} = \{\boldsymbol{\rho}^{(k)}\}_{1 \leq k \leq N}$. An example of event sequences is represented in Fig. 1(a) for $N = 3$. By definition, the observation sequence $\boldsymbol{O}$=$\boldsymbol{O}_1\boldsymbol{O}_2 \dots \boldsymbol{O}_U$ of an HMMESM results from the combination of the event sequences $\boldsymbol{E}$ with a valid scenario, $\mathscr{S}$, defined by:

$$\mathscr{S} = \{(\tau_u, \boldsymbol{s}_u)\}_{1 \leq u \leq U} \tag{1}$$

A scenario explains the detected event sequences $\boldsymbol{E}$. The term U represents the number of deep events. τ_u denotes the occurrence time of the u-th deep event whereas $\boldsymbol{s}_u$ is the associated signature. A signature $\boldsymbol{s}_u$ is the observable counterpart of a deep event. It is composed of one shallow event by observation channel. Shallow events can be either detected (true positive events: *tpe*s) or not (missing events: *me*s). A *me* is due to the misdetection of an event of interest in an observation channel. For example, the signature $\boldsymbol{s}_3$ of the Fig. 1(b) is made of two *me*s (black points) and of one *tpe* (vertical line).

In the sequel, $\tau_u^{(k)}$ denotes the occurrence time of the event associated with the u-th deep event and observed on the k-th observation channel and $m_u^{(k)}$ denotes the related measure. This measure is empty ($m_u^{(k)} = \emptyset$) if the associated event is a *me* ($\tau_u^{(k)} \notin \boldsymbol{\rho}^{(k)}$) and this measure is not empty ($m_u^{(k)} \neq \emptyset$) if the event is a *tpe* ($\tau_u^{(k)} \in \boldsymbol{\rho}^{(k)}$). The set $\boldsymbol{s}_u = \{\tau_u^{(k)}, m_u^{(k)}; 1 \leq k \leq N\}$ is referred to as the *signature* of the u-th deep event.

As shown in Fig. 1(b), a scenario defines a particular temporal cutting of $\boldsymbol{E}$ across all event sequences. Many scenarios may explain the detected events. However, $\mathscr{S}$ is considered as a *valid* scenario of $\boldsymbol{E}$ if it verifies:

$$\begin{cases} (i) & \tau_u, \tau_u^{(k)} \in I \\ (ii) & \tau_1 = \tau_1^{(k)} = 0, \tau_U = \tau_U^{(k)} = T \\ (iii) & u' > u \Rightarrow \tau_{u'} > \tau_u, \tau_{u'}^{(k)} > \tau_u^{(k)} \\ (iv) & \forall u, \exists k : m_u^{(k)} \neq \emptyset \end{cases} \tag{2}$$

Statement (ii) relates to the fictive signatures introduced at the beginning and the end of the sequences. Statement (iii) means that signatures preserve the ordering of their triggering deep events. Statement (iv) stipulates finally that one detected event at least is present per signature to fulfill the observability prerequisite of deep events.

In practice, the u-th observation of $\boldsymbol{O}$, $\boldsymbol{O}_u$, is made of the deep event occurrence time τ_u, of the associated signature $\boldsymbol{s}_u$, and of all the false positive events (*fpe*s) detected between $\boldsymbol{s}_{u-1}$ and $\boldsymbol{s}_u$, denoted $\boldsymbol{o}_{u-1,u}$, so that $\boldsymbol{O}_u = \{\tau_u, \boldsymbol{s}_u, \boldsymbol{o}_{u-1,u}\}$ except for $\boldsymbol{O}_1$, see Fig. 1(c).

2.2. Elements of an HMMESM

An HMMESM is completely specified by the parameter set $\boldsymbol{\lambda} = \{C, \boldsymbol{S}, \boldsymbol{A}, \boldsymbol{B}\}$ and by a random weight vector $\mathbf{w}$:

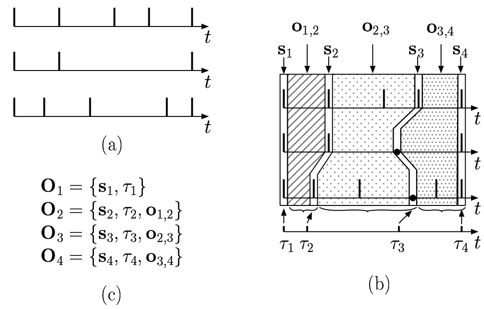

Fig. 1. Event sequences, scenario, observation sequence. (a) Set $\boldsymbol{E}$ of detected event sequences (N=3). (b) Example of a scenario $\mathscr{S}$ which explains $\boldsymbol{E}$. Black points composing the signatures correspond to missing events. (c) Constructing an observation sequence $\boldsymbol{O} = \boldsymbol{O}_1\boldsymbol{O}_2\boldsymbol{O}_3\boldsymbol{O}_4$ from a valid scenario $\mathscr{S}$ and a set $\boldsymbol{E}$ of detected event sequences.

• C denotes the cardinal of the state space $\boldsymbol{S} = \{S_j; 1 \leq j \leq C\}$ of the hidden process. S_1 and S_C are defined as the start and final states of the process. They account for the start and final signatures $\boldsymbol{s}_1$ and $\boldsymbol{s}_U$, respectively.

• $\boldsymbol{A} = \{a_{ij}\}$ is the transition probability matrix of the model, where $a_{ij} = P(X_u = S_j | X_{u-1} = S_i)$.

• $\boldsymbol{B} = \{b_{ij\mathbf{w}}(.|.)\}$ is the set of observation probability density functions (pdfs) attached to the hidden states of $\boldsymbol{S}$. We state:

$$\begin{aligned} &P(\boldsymbol{O}_u | \boldsymbol{O}_1^{u-1}, \boldsymbol{X}_1^u, \mathbf{W} = \mathbf{w}) = \\ &P(\boldsymbol{O}_u | \boldsymbol{O}_{u-1}, X_{u-1} = S_i, X_u = S_j, \mathbf{W} = \mathbf{w}) \\ &= b_{ij\mathbf{w}}(\boldsymbol{O}_u | \boldsymbol{O}_{u-1}) \end{aligned} \tag{3}$$

where $\boldsymbol{Z}_1^u$ stands for $Z_1 \dots Z_u$.

• $\mathbf{W}$ is a vector of N random variables,with realization $\mathbf{w} = (\mathrm{w}_1 \dots \mathrm{w}_N)^T$, $\sum_{k=1}^N \mathrm{w}_k = 1$, that reflects the confidence we have in each of the shallow processes.

2.3. Inference with HMMESM

The three basic problems of HMMs [1] also arise within the HMMESM framework, but with a somewhat different formulation due to the partially observable aspect of the observation sequence $\boldsymbol{O}$. Indeed, for a given set $\boldsymbol{E}$ of detected event sequences, multiple observation sequences $\boldsymbol{O}$ are possible. They have to be taken into account when solving the inference problem:

Evaluation: given $\boldsymbol{E}$, $\boldsymbol{\lambda}$ and $\mathbf{w}$, compute the likelihood: $P(\boldsymbol{E}|\boldsymbol{\lambda}, \mathbf{w}) = \sum_{\boldsymbol{O} \in \mathscr{S}(\boldsymbol{E})} \sum_{\boldsymbol{X}} P(\boldsymbol{O}, \boldsymbol{X} | \boldsymbol{\lambda}, \mathbf{w})$.

Decoding: given $\boldsymbol{E}$, $\boldsymbol{\lambda}$ and $\mathbf{w}$, infer the observation (the scenario) and state sequence that best explain $\boldsymbol{E}$: $\{\hat{\boldsymbol{O}}, \hat{\boldsymbol{X}}\} = \mathrm{argmax}_{\boldsymbol{O} \in \mathscr{S}(\boldsymbol{E}), \boldsymbol{X}} P(\boldsymbol{O}, \boldsymbol{X} | \boldsymbol{\lambda}, \mathbf{w})$.

Learning: given $\boldsymbol{E}$ and $\mathbf{w}$, adjust the parameter set $\boldsymbol{\lambda}$ to maximize the likelihood $P(\boldsymbol{E}|\boldsymbol{\lambda}, \mathbf{w})$.

$\boldsymbol{O} \in \mathscr{S}(\boldsymbol{E})$ denotes an observation sequence that results from the combination of $\boldsymbol{E}$ with a valid scenario $\mathscr{S}$ taken in the set $\mathscr{S}(\boldsymbol{E})$ of all valid scenarios that could explain $\boldsymbol{E}$. $\boldsymbol{X}$ denotes a possible state sequence at the origin of the observation sequence $\boldsymbol{O}$.

3. APPLICATION IN FMRI BRAIN MAPPING

Activation detection is formulated in terms of temporal alignment between sequences of hemodynamic response onsets (HROs) detected in the fMRI signal at voxel v and in the spatial neighborhood of v, and the input sequence of stimuli or stimulus onsets, whether an epoch- or an event-related paradigm is considered. This multiple event sequence alignment problem is solved within the probabilistic framework of HMMESMs. The deep process is the neural activation process that may take place at voxel v under stimulation (neural activation onset (NAO) events are the deep events). The related shallow processes are the fMRI signals observed at v and in its neighborhood (HRO events are the shallow events).

The benefit of estimating brain activity at voxel v by multiple event sequence alignment is fourfold. Robustness to noise can be achieved by taking into account spatial and temporal information directly into the core of the activation detection process. Analyzing only the non-stationarities of the fMRI signal rather than the whole signal makes it possible to model easier the timing function encoding the sequence of task-induced neural activations. It avoids hypothesizing about the shape of the hemodynamic response function (HRF) and about the linearity of the fMRI response.

The fMRI brain mapping procedure applied for each voxel v is depicted in Fig. 2.

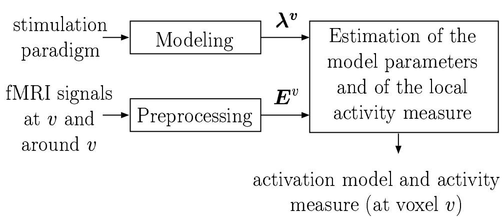

Fig. 2. Synoptic diagram of the brain mapping procedure at voxel v.

The set of multiple event sequences, $\boldsymbol{E}^v$, is obtained by preprocessing the fMRI signal at v and in its neighborhood. HRO event detection relies upon a wavelet decomposition of the fMRI signals [2]. For each voxel, 3 event sequences are derived using 3 different scales of decomposition. Therefore, when considering a V-connectivity 3D neighbourhood (V=26 in our application), the set $\boldsymbol{E}^v$ used to score neural activation at location v is made of 3(V+1) event sequences.

In the modeling step, the state space of the hidden process and the topology of the model are first defined from the stimulation paradigm. Let us consider a binary activation-baseline paradigm composed of P stimulation blocks. Among them, two are fictive blocks introduced at the beginning and the end of the paradigm for modeling purposes. Then, a state S_i, $1<i<C=P$, is used to model the i-th task-induced NAO event, or, equivalently, the i-th off-on paradigm transition. The start and final states S_1 and S_C are then added. A left-right topology is selected for the Markov chain modeling the deep process, with the additional constraints $a_{ii} = 0, \forall i$, and $a_{ij} = 0$ if $j>i+\Delta$, with $\Delta \sim \frac{C}{2}$, to prevent to declare active an fMRI signal that only responds to a few stimulation blocks. An example of Markov chain topology is depicted in Fig. 3 with Δ=2 and C=5. Concerning the state observa-

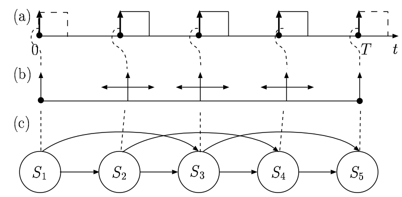

Fig. 3. Brain activation modeling. (a) input stimulation paradigm, (b) schematic representation of the task-induced NAO sequence, (c) resulting brain activation HMMESM.

tion pdf's, in order to model a possible activation lag l of the NAO events at voxel v, a random variable L is introduced in the expression of Eq. (3). Under some conditional independence assumptions of the observation components τ_u, $\boldsymbol{s}_u$, and $\boldsymbol{o}_{u-1,u}$, the expression of a state observation pdf can be factorized (Eq. (4)). In addition, we use data reduction techniques to limit the dimensionality of the observation.

$$\begin{aligned}
&b_{ijl\mathbf{w}}(\boldsymbol{O}_u|\boldsymbol{O}_{u-1}) = \\
&P(\boldsymbol{O}_u|\boldsymbol{O}_{u-1}, X_{u-1} = S_i, X_u = S_j, \mathbf{W} = \mathbf{w}, L = l) = \\
&P(\tau_u|X_u = S_j, L = l).P(\boldsymbol{s}_u|X_u = S_j, \tau_u) \\
&\;.P(\boldsymbol{o}_{u-1,u}|X_{u-1} = S_i, X_u = S_j, \boldsymbol{s}_u, \boldsymbol{s}_{u-1}) \qquad (4)
\end{aligned}$$

The first pdf in the product of Eq. (4) is a one-dimensional Gaussian defined as $\mathcal{N}(t_j + l, \sigma_j)$, where t_j denotes the time instant of the j-th off-on paradigm transition and l is the lag. The two other terms are modeled by Gaussian pdfs.

Finally, learning and mapping are performed at each location v based on the learning set $\boldsymbol{E}^v$. For processing time reasons, the learning problem is solved with a segmental k-Means algorithm [1] by maximizing the following criterion: $p^{\star,v} = max_{\boldsymbol{O}\in\mathscr{S}(\boldsymbol{E}^v),\boldsymbol{X},l,\boldsymbol{\lambda},\mathbf{w}} P[\boldsymbol{O}, \boldsymbol{X}|l, \boldsymbol{\lambda}, \mathbf{w}]$. Mapping is finally performed based on the likelihood map $\{p^{\star,v}\}$ initially transformed into a p-value map [2].

4. RESULTS

The HMMESM-based brain mapping method was applied to synthetic and real epoch-related fMRI data. Five epoch-related synthetic data sets (DS1–DS5) were derived from a single real noise fMRI data set (no activity) using fake activation patterns embedded at known locations. The data sets DS1–DS3 were designed to illustrate the signal to noise variability. The activation pattern was obtained by convolving the HRF model proposed by the statistical parametric mapping (SPM) software [4] with the "expected" boxcar-like timing function (150 scan length, 7 blocks ON interleaved with 8 blocks 0FF, 10 scans each). The SNR associated with DS1, DS2 and DS3

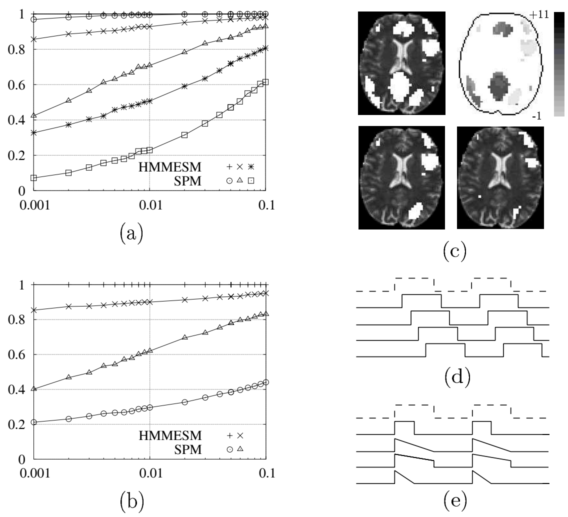

Fig. 4. (a,b) Detection results (ROC curves) on epoch-related synthetic data for (a) DS1 (‘+’,‘○’), DS2 (‘×’,‘△’) and DS3 (‘∗’,‘□’), and for (b) DS4 (‘+’,‘○’) and DS5 (‘×’,‘△’). (c) Results on epoch-related real data. Activated brain regions are represented in white. Top left: HMMESM map. Top right: HMMESM activation lag map. A gray level scale graduated in number of scans is used to represent the local activation lag estimate. Bottom left: HMMESM in-phase activation map. Bottom right: SPM map. (d,e) “expected” boxcar-like timing function (dotted line) and activation patterns used to build the synthetic data sets DS4 and DS5 (thin line), see text.

was large, medium and small, respectively (following the definition of [3], SNR=1, 0.5 and 0.25, respectively). The data set DS4 was designed to illustrate timing variability between stimulation paradigm and neural/hemodynamic response. Four activation patterns were obtained by convolving the HRF of the SPM sotware with four delayed versions of the “expected” boxcar-like timing function (see Fig. 4(d), only 2 ON blocks are presented). The last data set DS5 was designed to illustrate variability in the shape of the hemodynamic response to a stimulation block (unsustained activation during task). Four noise-free activation patterns were computed by convolving the HRF of the SPM sotware with the four deterministic functions plotted in Fig. 4(e). Finally, 42 real epoch-related fMRI data sets were acquired from healthy subjects who were asked to perform auditory lexical processing. The SPM method was employed as the main comparator of the HMMESM method (SPM2 release). It makes use of a single regressor, namely, the HRF model proposed by the software convolved with the “expected” boxcar-like timing function. The three synthetic data sets DS1–DS3 should rather be considered as an “ideal” application case of SPM since the regressor used for SPM analysis completely matches the noise-free activation pattern to be detected.

With respect to the synthetic data sets, activation detection performance of the HMMESM and SPM methods are compared from receiver-operating characteristic (ROC) curves. Results obtained for DS1,DS2 and DS3 demonstrate the robustness to noise of the HMMESM method when compared to SPM (see Fig. 4(a)). They can largely be explained by the multiple-event-sequence-based, spatial information fusion strategy used to detect local neural activity. Results obtained for DS4 show clearly the complete insensitivity of the HMMESM method to timing variations of the hemodynamic response with respect to the input stimulation blocked paradigm (see Fig. 4(b)). They point out the contribution of a statistical modeling of the activation delay. Results obtained for DS5 demonstrate the relative insensitivity of the HMMESM method to variations of the hemodynamic response to a stimulation block (see Fig. 4(b)). These results validate the strategy that consists in focusing the analysis on the transient events that are the HROs.

With respect to the real epoch-related data sets, activation detection results obtained by the HMMESM and SPM methods have been compared based on activation maps. HMMESM maps are in very good accordance with SPM ones. Besides, the HMMESM maps show additional *delayed* activation areas with respect to the stimulation blocked paradigm. A representative example of HMMESM and SPM mapping results is shown in Fig. 4(c).

5. CONCLUSION

A markovian model dedicated to the analysis of multiple event sequence based random processes has been presented. From this model, an unsupervised fMRI brain mapping method has been developed. By accounting for spatial information within a statistical framework of multiple event sequence detection, multiple event sequence fusion, the HMMESM based mapping method shows high robustness to noise and variability of the active fMRI signal.

6. REFERENCES

[1] L.R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *In Proc. IEEE*, vol. 77, pp. 257–286, 1989.

[2] S. Faisan, L. Thoraval, J.-P. Armspach, M.N. Metz-Lutz, and F. Heitz, “Unsupervised learning and mapping of active brain functional MRI signals based on hidden semi-Markov event sequence models,” *IEEE Transactions on Medical Imaging*, vol. 24, no. 2, pp. 263–276, 2005.

[3] N. Lange, S.C. Strother, J.R. Anderson, F.A. Nielsen, A.P. Holmes, T. Kolenda, R. Savoy, and L.K. Hansen, “Plurality and resemblance in fMRI data analysis,” *NeuroImage*, vol. 10, pp. 282–303, 1999.

[4] K.J. Friston, J. Ashburner, and al, “SPM 97 course notes,” *Wellcome Department of Cognitive Neurology, University College London*, vol. 13, pp. 13–25, 1997.