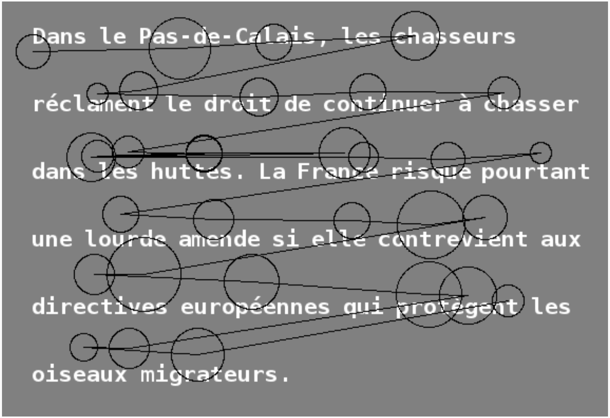
Joint eye-movement and EEG analysis using Hidden semi-Markov Models to identify and characterize reading strategies

Brice Olivier

19th day of AIGM network, December 5, 2018

Supervisors : Jean-Baptiste Durand, Anne Guérin-Dugué
Laboratories : `Mistis, Inria / Laboratoire Jean Kuntzmann & Vibs, Gipsa-lab`
Funding : LabEx PERSYVAL-Lab

Eye-tracking vocabulary



Fixation (circles): immobilization of visual gaze during few ms.
Saccade (lines): brief movement of the eye between two fixation.
Scanpath: series of fixations and saccades recorded during a given task.

Previous studies on reading processes

- Eye-movement contains information in the reading processes
- Precursors analyzed **microprocesses of reading**. e.g. longer fixations with misspelled words / less common words, (Rayner, 1998, Hyrskykari et al. 2000, 2003)
- We focus on reading as a whole mechanism (scanpath) in the context of **information search** tasks involving both **semantic information gathering** and **decision making processes**, (Frey et al. 2013).
- Reading strategies are used as **gears** for different purposes, (Carver, 1990)

State of the art on data-driven studies and problematic

- Reading strategies discovery in information search task using **Hidden Markov Models** (Simola et al., 2008)

- How to segment scanpaths into interpretable zones, in terms of cognitive phases in information acquisition and processing ?

- Hypothesis: reading strategies changes reflect cognitive steps, expressed indirectly through eye movements regime changes (and Markovian regime changes)

Issues:

- HMM have non-fitted sojourn distributions for reading strategies –> HSMM
- Legitimacy of the ”cognitive processes” using eye-movement data only –> Couple EM&EEG data

Material and Method - Goal and experimental settings

Simulate press review task through binary decision:

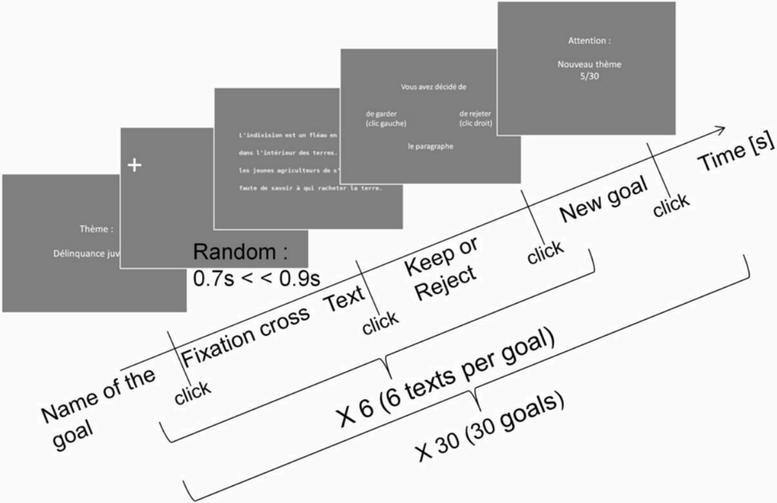
Is the text related to the topic or not ?

- positive decision: target words
- negative decision: incongruent words

Experimental settings

- 15 participants
- 180 texts per participants (Extracted from the french newspaper LeMonde)
- Goal is a nominal phrase. e.g. ”modern art”
- 60 Highly / 60 Moderately / 60 Un - related texts to the topic (Latent Semantic Analysis control)

Material and Method - Experimental Procedure



Frey et al. 2013

Outline

Hidden (semi-)Markov Models

Analysis protocol on eye-movements using EEG as covariates

Joint modeling of eye-movements and EEGs

Conclusion & Perspectives

Quick reminder of a HMM - Model

Let $\{S_1, ..., S_T\}$, a sample of T R.V. corresponding to a discrete Markov Chain and taking values into a finite state space $\mathcal{S}$, with $Card(S)$=M.

Let $\pi_j \equiv P(S_1 = j)$, the probability that the first state is j. $\pi = \{\pi_j\}$ is a $1 \times M$ vector denoting **initial distribution**.

Using *Markov property*, let $a_{ij} \equiv P(S_t = j|S_{t-1} = i)$, the probabilities to transit from state i to j at time t. $A = \{a_{ij}\}$ is a $M \times M$ **transition matrix**.

Consider an output $\{O_1, ..., O_T\}$, $\forall t, O_t \in \mathcal{V}$ with $Card(\mathcal{V}) = K$, the observation space. O_t depends conditionally on S_t, so we have $b_j(v_k) \equiv P(O_t = v_k|S_t = j)$, the **Observation conditional probabilities**.

Quick reminder of HMM - Inference and Learning

Inference

Forward-Backward algorithm to maintain tractability in inference problems (e.g. probabilities such as $P(o_1, ..., o_T|\theta)$)

Learning

MLE of $\hat{\theta}$ via the iterative **Expectation-Maximization** algorithm:

E-step: decomposition of $E[\ln P(S_{1:T}, O_{1:T}|\theta)|O_{1:T}, \theta^{old}]$ and simplification

M-step: maximize with respect to θ

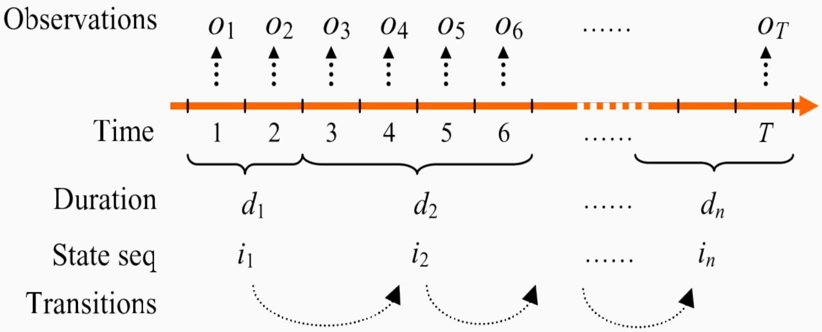
Restoration

Estimation of the most likely state sequence using the **Viterbi** algorithm:

$$\hat{S}_{1:T} = \underset{S_{1:T}}{\arg\max} \; P(S_{1:T}|O_{1:T}, \hat{\theta})$$

Definition of a HSMM

Yu, 2010



State duration $\forall t, D_t \in \mathcal{D} = \{1, \ldots, \infty\}$

Sojourn distribution: $p_j(d) \equiv P(D_t = d, S_{t+1:t+d} = j, S_{t+d+1} \neq j | S_t = j)$

Model parameters: $\theta \equiv \{a_{ij}, b_j(v_k), \pi_j, p_j(d)\}$

Outline

Hidden (semi-)Markov Models

Analysis protocol on eye-movements using EEG as covariates

Joint modeling of eye-movements and EEGs

Conclusion & Perspectives

Output process construction

Observed Process: ”Readmode”

Categorial variable with 5 levels from long regression to long progression, i.e. **bounded number of identified words in one saccade** $\in \{< -1, -1, 0, 1, > 1\}$

Time index: fixation onset

Latent Process

Reading strategies ”hidden”, observed through the readmode

Number of reading strategies unknown and to be determined

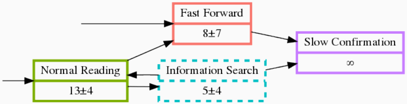
Strong hypothesis

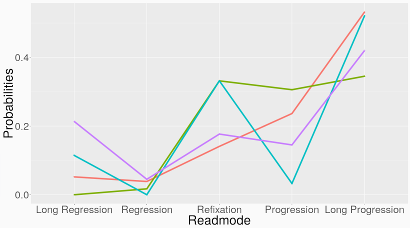
Scanpaths are independant and identically distributed

Model covariates

Fixation duration, Saccade amplitude, Textual properties, EEGs

Estimated model parameters

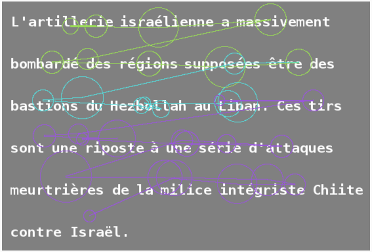




Each reading strategy is characterized by: a **readmode pattern**, a **sojourn distribution**, **probabilities to transit** to other reading strategies and an **initial probability**

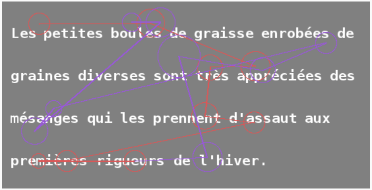
Scanpath Segmentation (restoration of state sequences)

"Israeli-Palestinian conflict" (Highly Related)



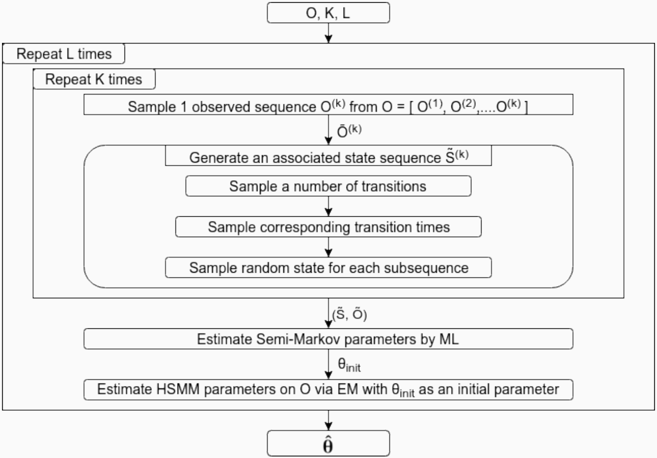
"Israeli artillery massively bombed areas believed to be Hezbollah strongholds in Lebanon. These shootings are a response to a series of deadly attacks by the fundamentalist Shiite militia of Israel."

"Birds hunt" (Moderately Related)



"Small balls of fat coated with various seeds are much appreciated by the tits that storm them at the first severe winter."

Model selection - Random Initialization for HSMM



Algorithm: Random initializations for H(S)MM

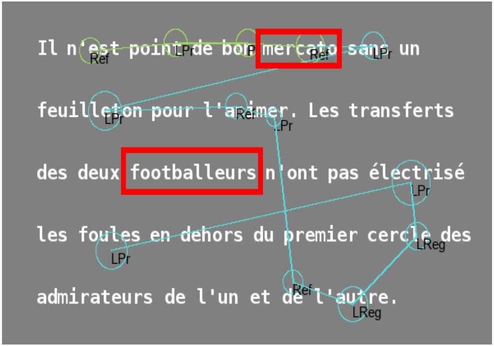
Model selection - Choosing number of Classes

#Classes	LL	BIC	Entropy	ICL*	#Params
3	-50835	-101873	11326	-124525	20
4	-50726	-101814	**9753**	**-121321**	26
5	-50645	-101651	11886	-125423	34
6	**-50593**	**-101556**	14748	-131054	35

* Integrated Complete Likelihood is equivalent to BIC with an additional penalty term: the conditional entropy of the hidden variable (Biernacki et al., 2000, Volant et al., 2012)

Model validation - Text - Meta Word Representations

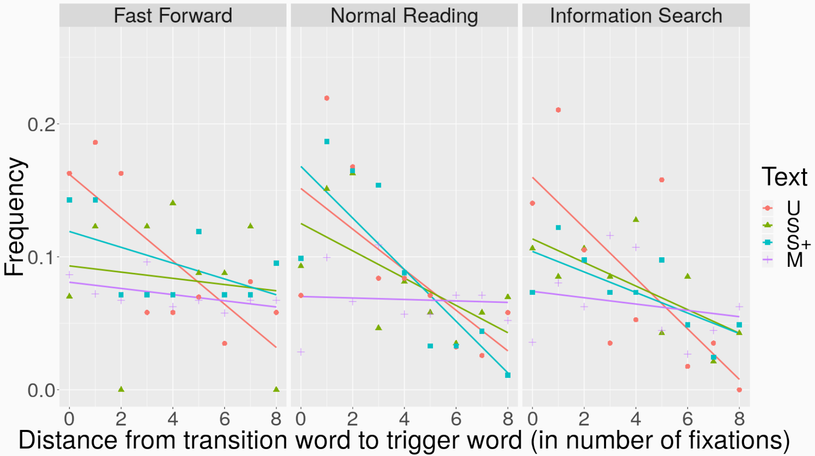
"Russia's chief" - Unrelated



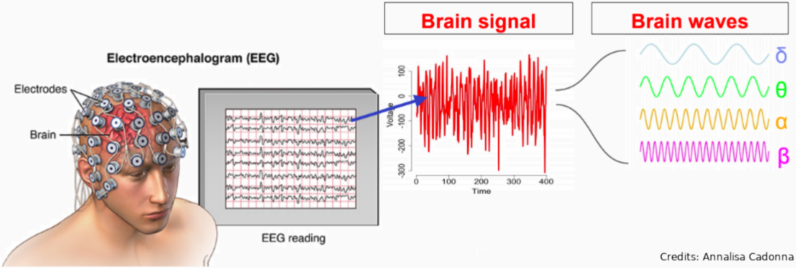
"There is no good mercato without a soap opera to animate it. The transfers of the two footballers did not electrify the crowds outside the first circle of admirers of the one and the other."

- Automatic detection of trigger words w.r.t. topics
- We used Facebook's fastText word representations and combined them into a single framework with adjusted cosine similarity measure
- $w_{target} = \arg\max_i \frac{w_{topic} \cdot w_{text,i}}{\|w_{topic}\| \|w_{text,i}\|}$ where $w_.$ is a vector word representation and $w_{text,i}$ stands for i-th word in text.

Model validation - Text - distance of target words to transitions



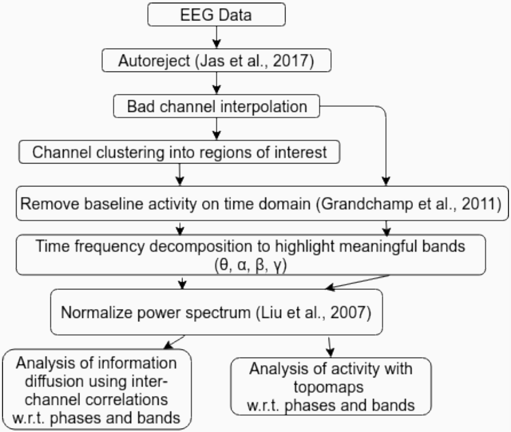
Model validation - EEGs - Bands, activities, tasks

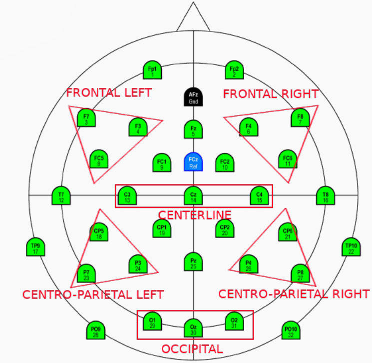


Credits: Annalisa Cadonna

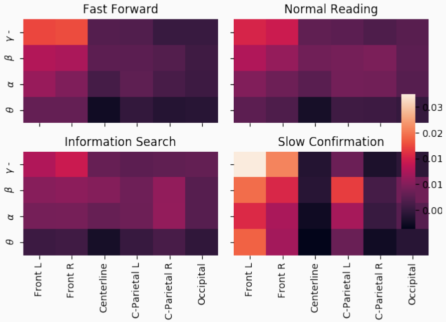
- α and θ oscillations reflect memory performances (Klimesch, 1998),
- "Memory is an extremely distributed system with long term memory located primarily in posterior cortices and accessed from prefontal regions" (Klimesch et al., 2005, 2006, 2011)
- Memory encoding and restitution differences observed in α band (Seidkhani et al., 2017)

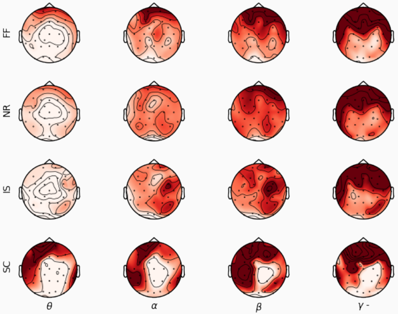
Model validation - EEGs - Analysis Methodology



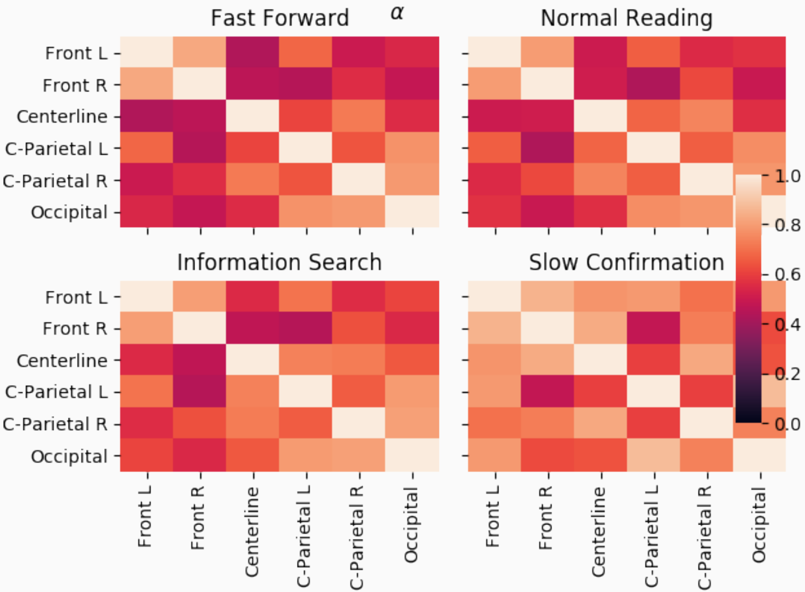


Model validation - EEGs - Activity





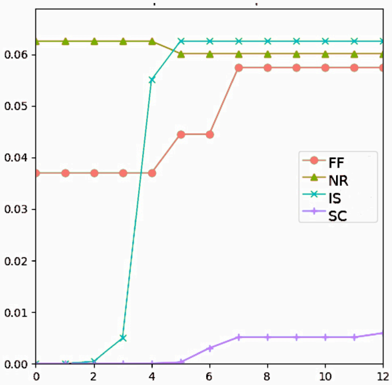
Model validation - EEGs - Information Diffusion



Issues - Uncertainty of state sequence restoration

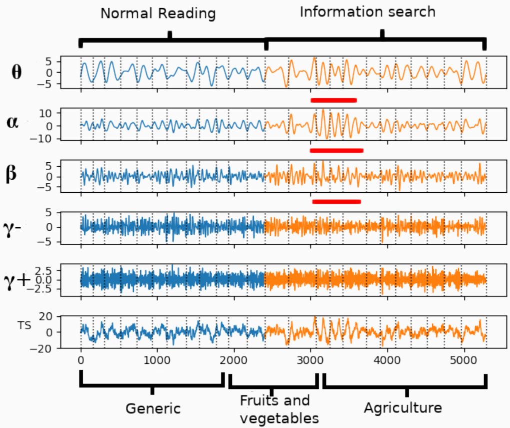
- State Entropy computation (Durand et al. 2012) has shown a high amount of uncertainty.
- Computation of Posterior probabilities of state sequence:

$$s_t^{(k)} = \max_{s_{1:t-1}, s_{t+1:T}} P(S_{1:t-1} = s_{1:t-1}, s_t = k, S_{t+1:T} = s_{t+1:T} | O_{1:T})$$



Issues - delay in EEGs

"Economic growth" - Unrelated text



- Visible delay between ocular and brain activities
- "Eye movements provide information to the brain which guides the eyes in return.", (Frey et al. 2013)

Outline

Hidden (semi-)Markov Models

Analysis protocol on eye-movements using EEG as covariates

Joint modeling of eye-movements and EEGs

Conclusion & Perspectives

Modeling specifications

Different sampling rates

- $t \in \{1, ..., T\}$ now denotes a temporal index in ms.
- Let N_t, the number of fixations from 1 to t

Delayed State

- Let $\{S_1^{(2)}, ..., S_T^{(2)}\}$ a discrete latent state taking values in $\mathcal{S}$ and encoding the first SMC $\{S_1, ..., S_{N_T}\}$ at a higher sampling rate, plus a lag.
- We denote the lag $\{\epsilon_{N_1}, ..., \epsilon_{N_T}\}$, with $\epsilon_{N_t} \in \{1, ..., \mathcal{L}\}$ in its most general form.
- Hence we have: $S_{t+\epsilon_{N_t}}^{(2)} = S_{N_t}$
- ϵ_{N_t} could be deterministic, random, autoregressive, conditional to channels or states.

Modeling EEG output process

Output process related to EEGs models wavelet coefficients

Single output process

- $O_t^{(2)}$ is a concatenation of all wavelet coefficients and all channels at time t.
- $P(O_t^{(2)}|S_t^{(2)} = j) = \mathcal{N}(0, \Sigma_j)$
- Invariant regarding time since it is already encapsulated in $S_t^{(2)}$
- Sparse covariance matrix

Inference differences wrt HSMM

With a hypothesis s.t. $(\epsilon_1, ..., \epsilon_T)$ is a SMC, two sets of parameters are introduced.

In E-step:

- most of the sufficient statistics remain the same but with an additional loop over $\mathcal{L}$, the delay,
- a new sufficient statistic is introduced:

$$P(\epsilon_{N_t} = l, \epsilon_{N_{t-1}} = l', S_{N_t} = k, F_{N_{t-1}} = 1 | O_{N_1}, ..., O_{N_T}, O_1^{(2)}, ..., O_T^{(2)})$$

  where $F_{N_{t-1}} = 1$ denotes the fact that a transition will occur at time N_t.

Inference algorithm in $O(T\mathcal{L}M^2D)$ vs $O(TM^2D)$ for HSMM vs $O(TM^2)$ for HMM...

Outline

Summary

- We provided a full statistical analysis methodology for analyzing complex signals,
- we dug deeper in the understanding of reading mechanisms in press review-like tasks,
- we proposed to assemble asynchronous and heterogeneous signals in a single probabilistic model.

Perspectives

- **Random parameter initialization algorithm for HSMM:** provide heuristics for the choice of hyperparameters to increase speed
- **EEGs:** Hypothesis testing on significant inter-channel correlations –> Graph representations
- **EEGs:** investigate inter-channel partial correlations
- **Model testing:** Get first results on simulations for the proposed model to ensure identifiability
- **Scaling:** new model highly complex + high usage of RAM to store MODWT coefficients of EEGs –> Online or preferably minibatch version.

References i

Aline Frey, Gelu Ionescu, Benoit Lemaire, Francisco López-Orozco, Thierry Baccino, and Anne Guérin-Dugué, *Decision-making in information seeking on texts: an eye-fixation-related potentials investigation.*, Frontiers in systems neuroscience **7** (2013), no. August, 1–22.

Yann Guédon, *Estimating Hidden Semi-Markov Chains From Discrete Sequences*, Journal of Computational and Graphical Statistics **12** (2003), no. 3, 604–639.

Jaana Simola, Jarkko Salojärvi, and Ilpo Kojo, *Using hidden Markov model to uncover processing states from eye movements in information search tasks*, Cognitive Systems Research **9** (2008), no. 4, 237–251.

Shun Zheng Yu, *Hidden semi-Markov models*, 2010, pp. 215–243.

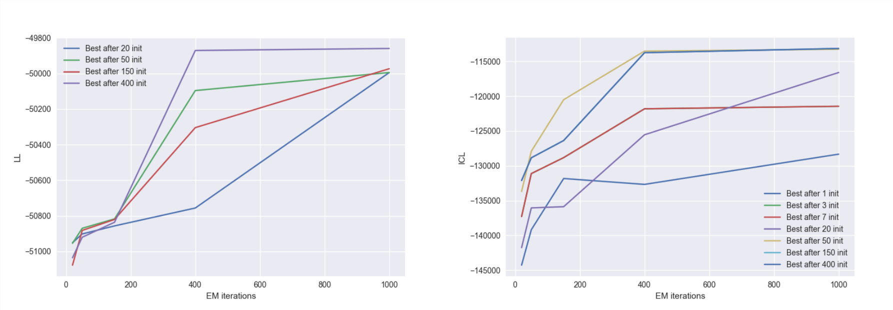
Outline

Supplementary material

Quick reminder of HMM - Limitations

- State sojourn time are by definition Geometric
- Let $X \sim G(p), \mathbb{E}[X] = 1/p, \mathbb{V}[X] = \frac{1-p}{p^2}$. Expectation and Variance linked by one single parameter p.

Model selection - RandomInit - choosing of K, L



Model validation - Several indicators

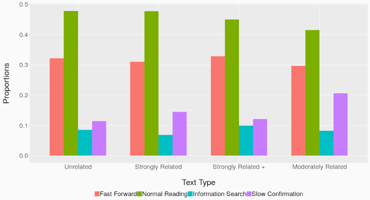
		Normal reading	Fast Forward	Information Search	Slow Confirmation
Fixation duration (ms)		183 ± 68	170 ± 60	190 ± 70	188 ± 68
Saccade amplitude (px)		121 ± 103	150 ± 94	136 ± 103	144 ± 98
Reading speed (wpm)		382	600	436	227
Cumulated cosine*		$.33 \pm .28$	$.33 \pm .30$	$.51 \pm .23$	$.47 \pm .26$
Saccade direction	Backward	.09	.09	.18	.19
	Upward	.01	.02	.04	.10
	Downward	.14	.22	.19	.19
	Forward	.71	.61	.51	.44
	Last	.05	.05	.07	.08

* Measure of cumulated gathered semantic information

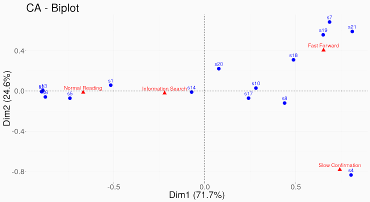
Speed reading suggests to be an easy task and therefore shorter fixations - *Rayner (1998), Simola et al. (2008)*

Model validation - Understanding the usage

Strategies usage wrt text types



Factorial Correspondence Analysis: Strategies and Subjects



In practise:

- strategies are used differently according to the text type,
- not all strategies are used for every trial or by every subject.

Model validation - EEGs - Information Diffusion