Joint analysis of eye-movement and EEGs using coupled hidden semi-Markov models to identify and characterize reading strategies

Brice Olivier

Supervisors : Jean-Baptiste Durand, Anne Guérin-Dugué

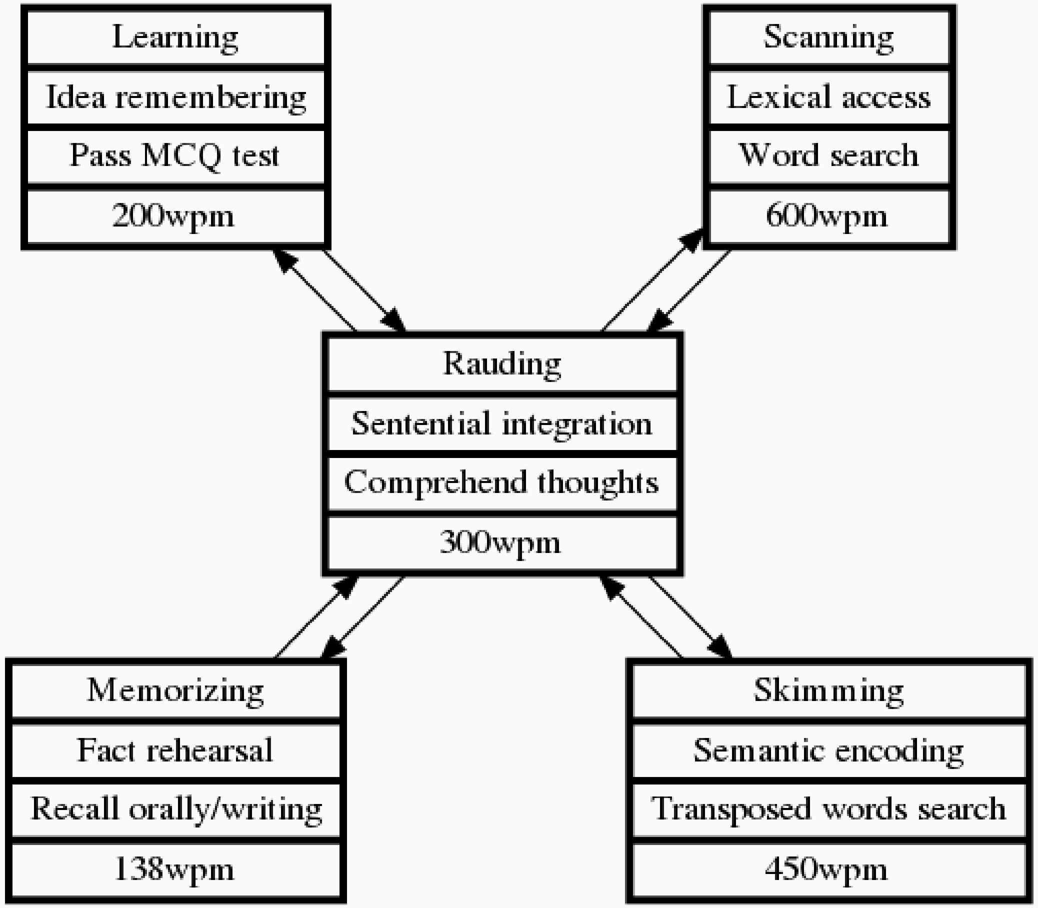
University Grenoble Alps - MSTII doctoral school
Laboratories : `Mistis, Inria / Laboratoire Jean Kuntzmann & Vibs, Gipsa-lab`

Funding : LabEx PERSYVAL-Lab

PhD Defense, June 26, 2019

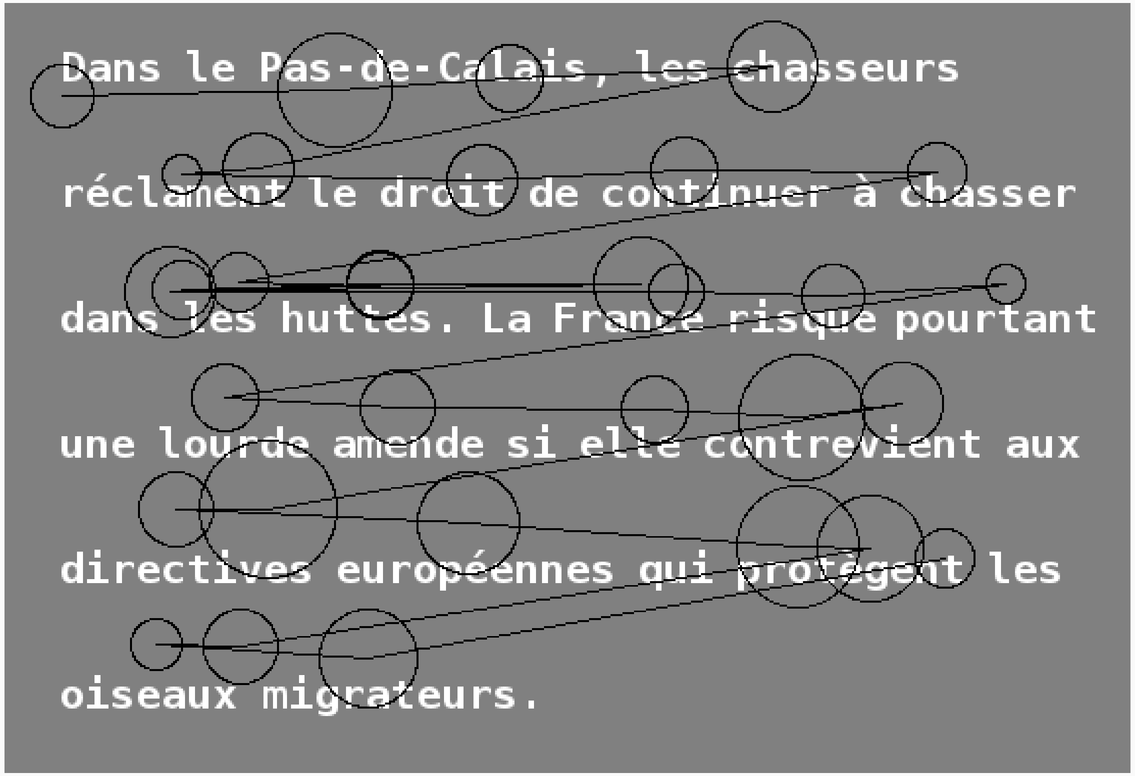
Reading strategies & segmentation (Carver, 1990)

- Reading strategies are used as **gears** to achieve different goals
- Each goal require different **cognitive components** and has different outcomes
- Carver, 1990 identified **5 reading strategies**
- Reading strategies may be discriminated using **reading rate**
- proficient readers are not faster but changes gear more efficiently
- **Issue**: controled protocols to isolate a single reading strategy



Eye-movement & Eye-tracking

- During reading, eyes move across words and can be tracked with an **eye-tracker**



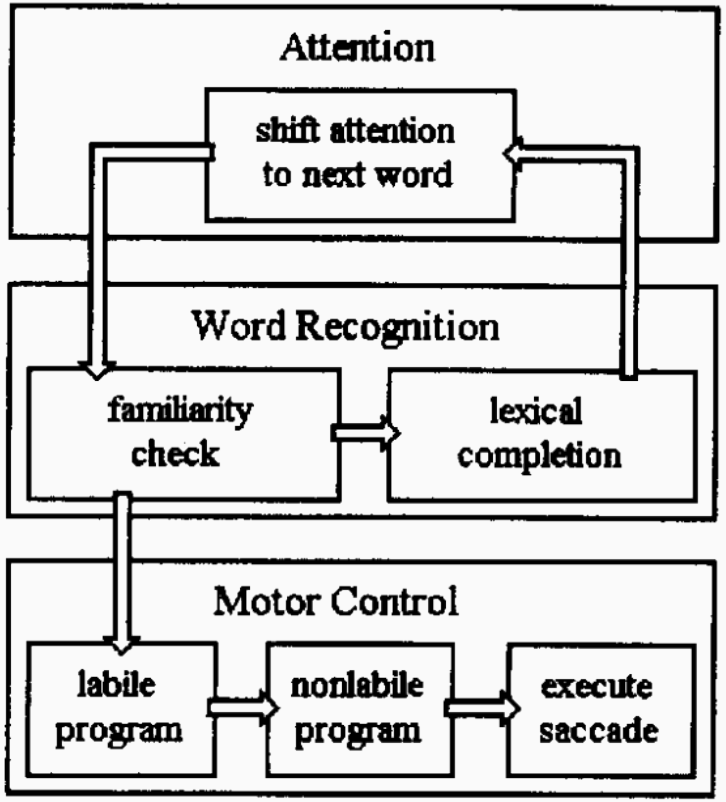
- **Fixation** (circles): immobilization of visual gaze during few ms.
- **Saccade** (lines): brief movement of the eye between two fixations.
- **Scanpath**: series of fixations and saccades recorded during a given task.

Microprocesses of reading

- 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)

- E-Z reader (Reichle et al., 1998)
  - model for each word
  - rule-based model
  - Output: probability to fix a next word
- **Issue**: does not take into account the heterogeneity of the task



Material and Method - Ecological settings (Frey et al., 2013)

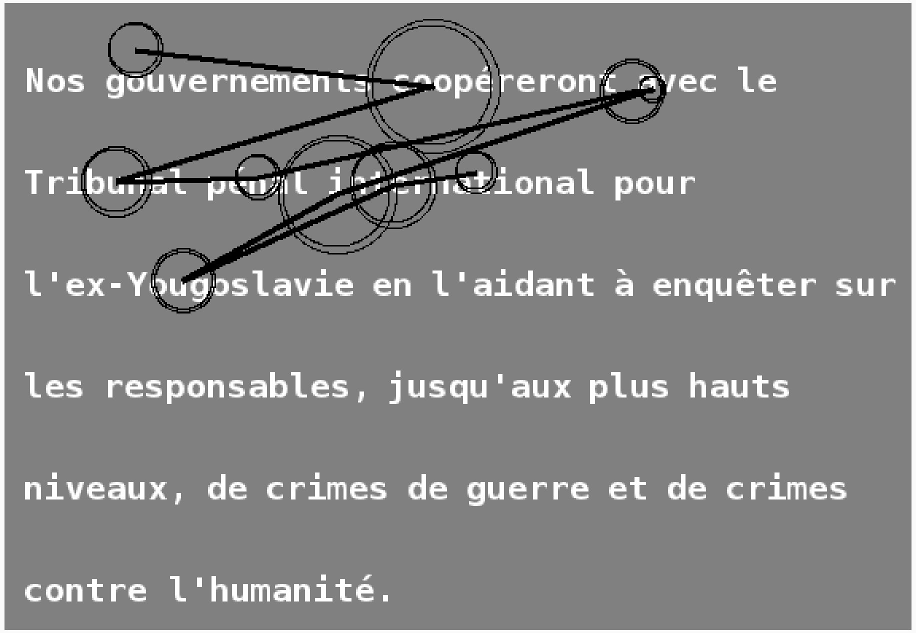
- Ecological context: **information search** tasks involving both **semantic information gathering** and **decision making processes** (Frey et al., 2013).
- Simulate press review task through binary decision:

  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)
- Datasets: eye-movement, electroencephalograms (EEGs), behaviors

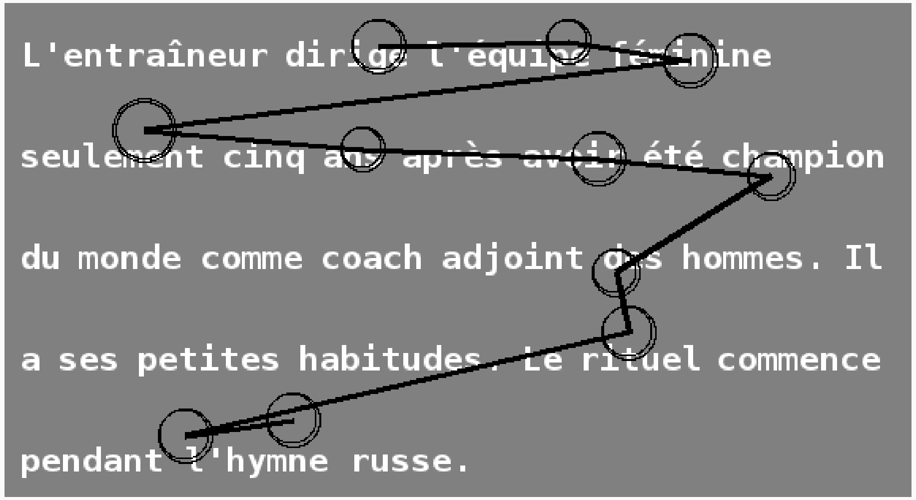
Scanpath examples - HR / UR texts

"International tribunal" (Highly Related)



"The coach leads the women's team only five years after being world champion as assistant coach of men. He has his little habits. The ritual begins during the Russian anthem."

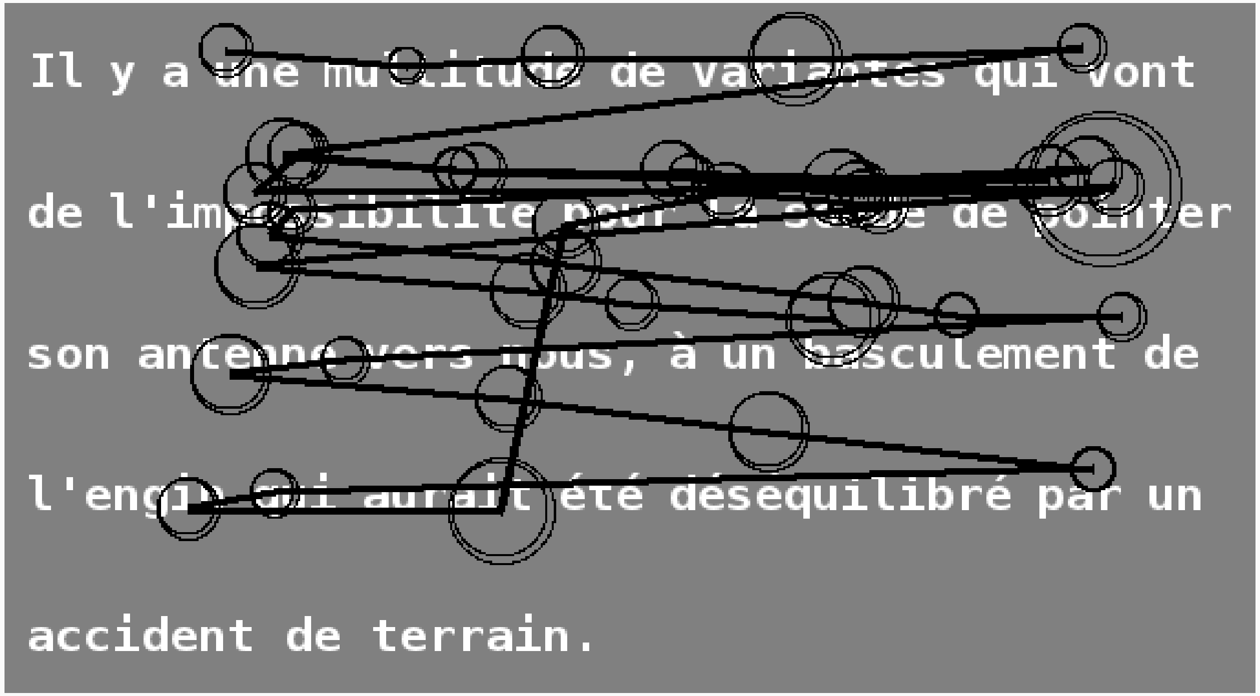
"Iraq conflict" (Unrelated related)



"The French football team defeated Australia in the World Cup final. The coach is very satisfied with his team and looks forward to future matches with enthusiasm and serenity."

Scanpath example - MR text

"Planets observation" (Moderately Related)



"There are a multitude of variants that range from the impossibility for the probe to point its antenna towards us, a tilting of the machine that would have been unbalanced by a terrain accident."

Data-driven scanpath segmentation

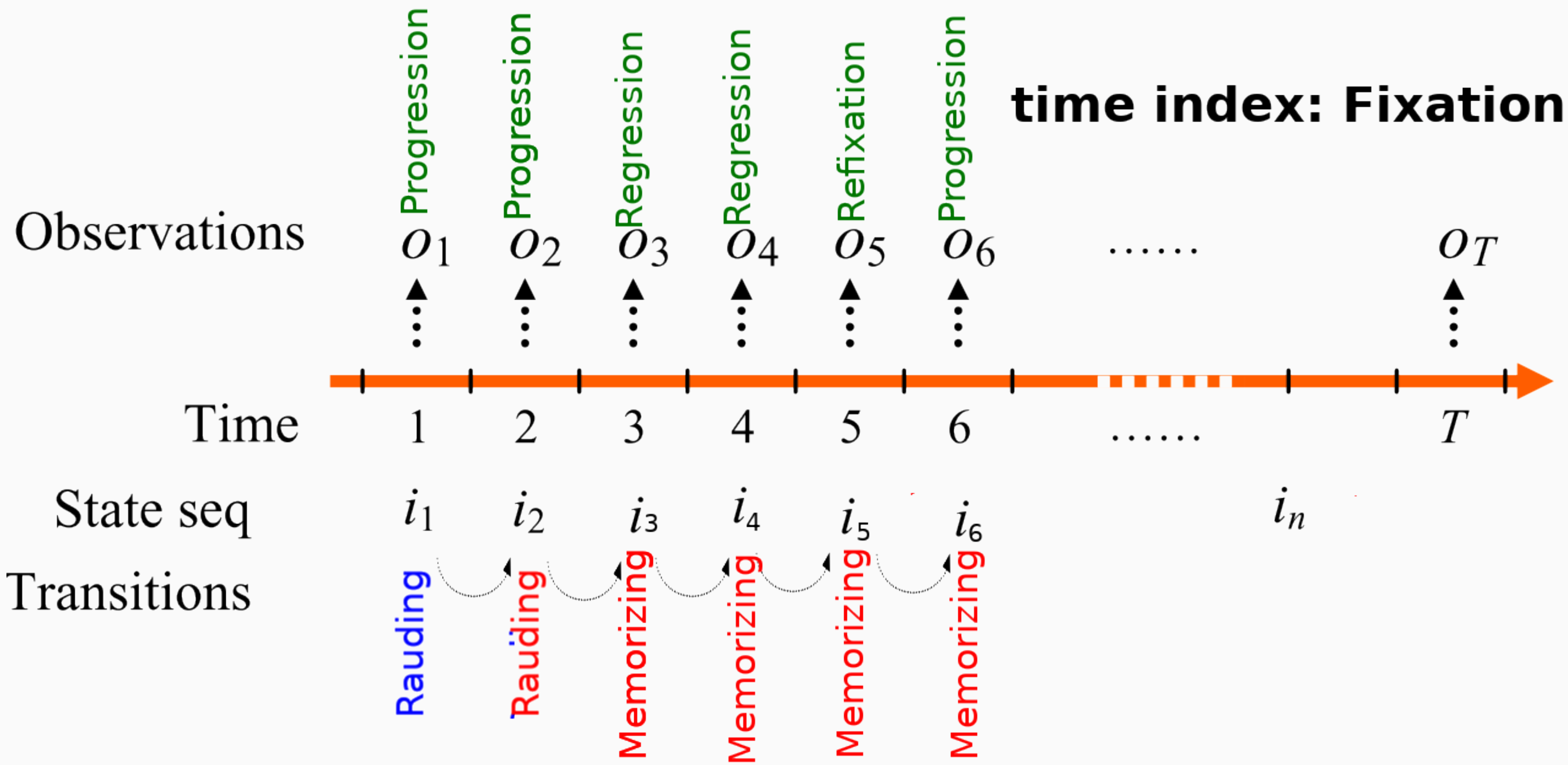
- Based on eye-movement features, **how to segment scanpaths** into intrepretable zones (reading strategies) which reflect changes in cognitive processes in information acquisition and processing ?
- How can we use **covariates** (Texts, EEGs) to reinforce interpretation of the segmentation based on eye-movement?
- How can we **model both eye-movement and EEGs** into a framework to enhance segmentation ?
- Need for a **statistical tool** to segment temporal data (Simola, Salojärvi, and Kojo, 2008)

Outline

1. Hidden (semi-)Markov Models
2. Initialization strategies for Expectation-Maximization
3. Analysis protocol on eye-movement
4. Segmentation a posteriori analysis of covariates (text, EEGs)
5. Joint modeling of eye-movement and EEGs
6. Contributions & Perspectives

Outline

Hidden Markov Model (HMM)



$\mathcal{S} = \{Rauding, Skimming, Memorizing\}$ $\mathcal{V} = \{Regression, Refixation, Progression\}$

State sequence $\mathbf{S} = \{S_1, ..., S_T\}$ Observation sequence $\mathbf{O} = \{O_1, ..., O_T\}$

Model parameters:

Initial probabilities

$\forall j \in \mathcal{S}, \quad \pi_j = P(S_1 = j)$

Transition probabilities

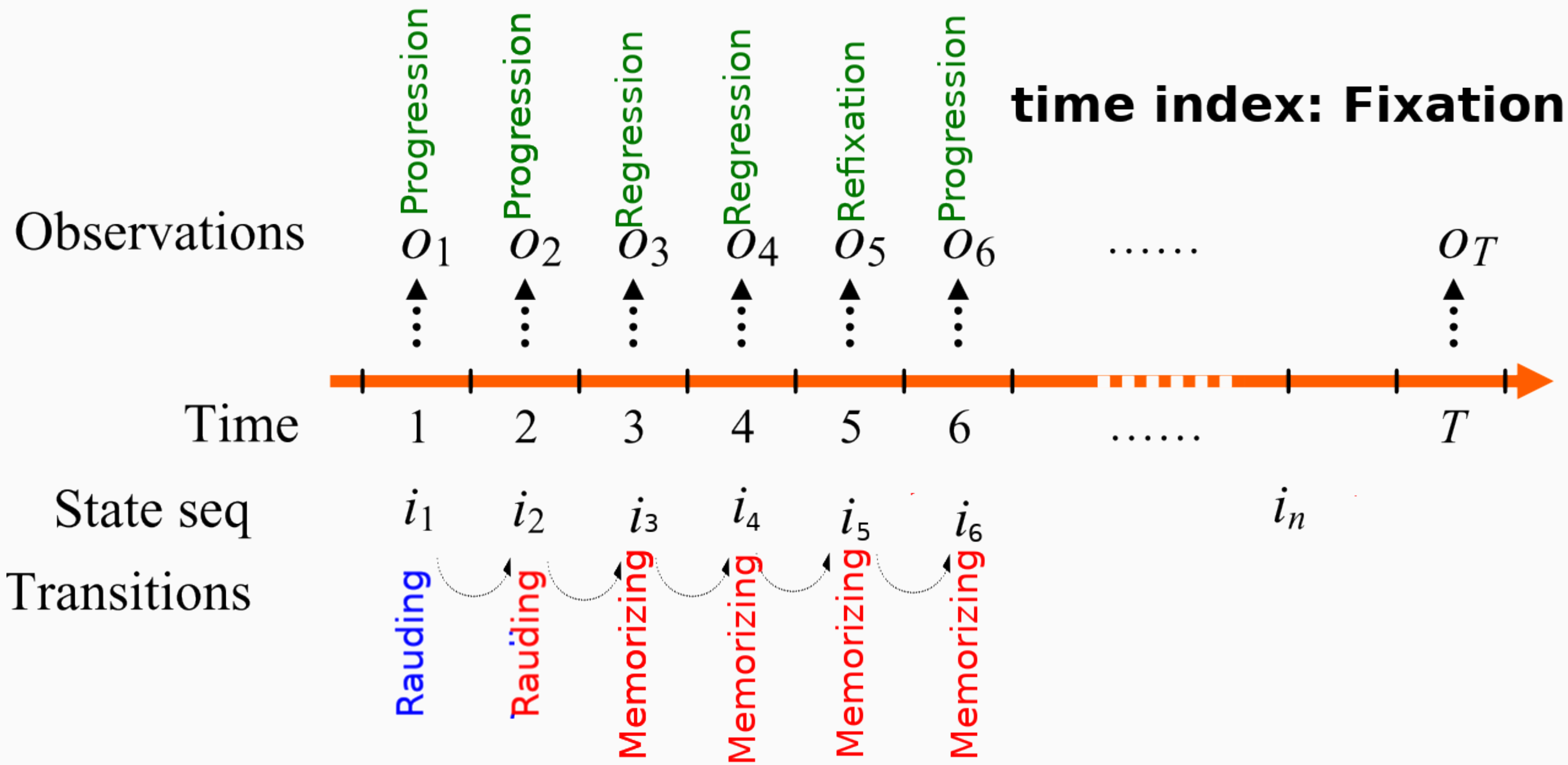
$\forall i, j \in \mathcal{S}, \quad A_{ij} = P(S_t = j | S_{t-1} = i)$

emission distributions

$\forall j \in \mathcal{S}, v_g \in \mathcal{V}$

$b_j(v_g) = P(O_t = v_g | S_t = j)$

Hidden Markov Model (HMM)



Say $\theta = \{\pi_j, A_{ij}, b_j(v_g)\}$ are unknown, **S** hidden, and I just observed **O**:

- How do I estimate the model parameters $\hat{\theta}$?

- How do I compute the most likely state sequence $\arg\max_{S_{1:T}} P(S_{1:T}|O_{1:T}, \hat{\theta})$?
- How do I find $Card(\mathcal{S})$?
- How do I identify $\mathcal{S}$?

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

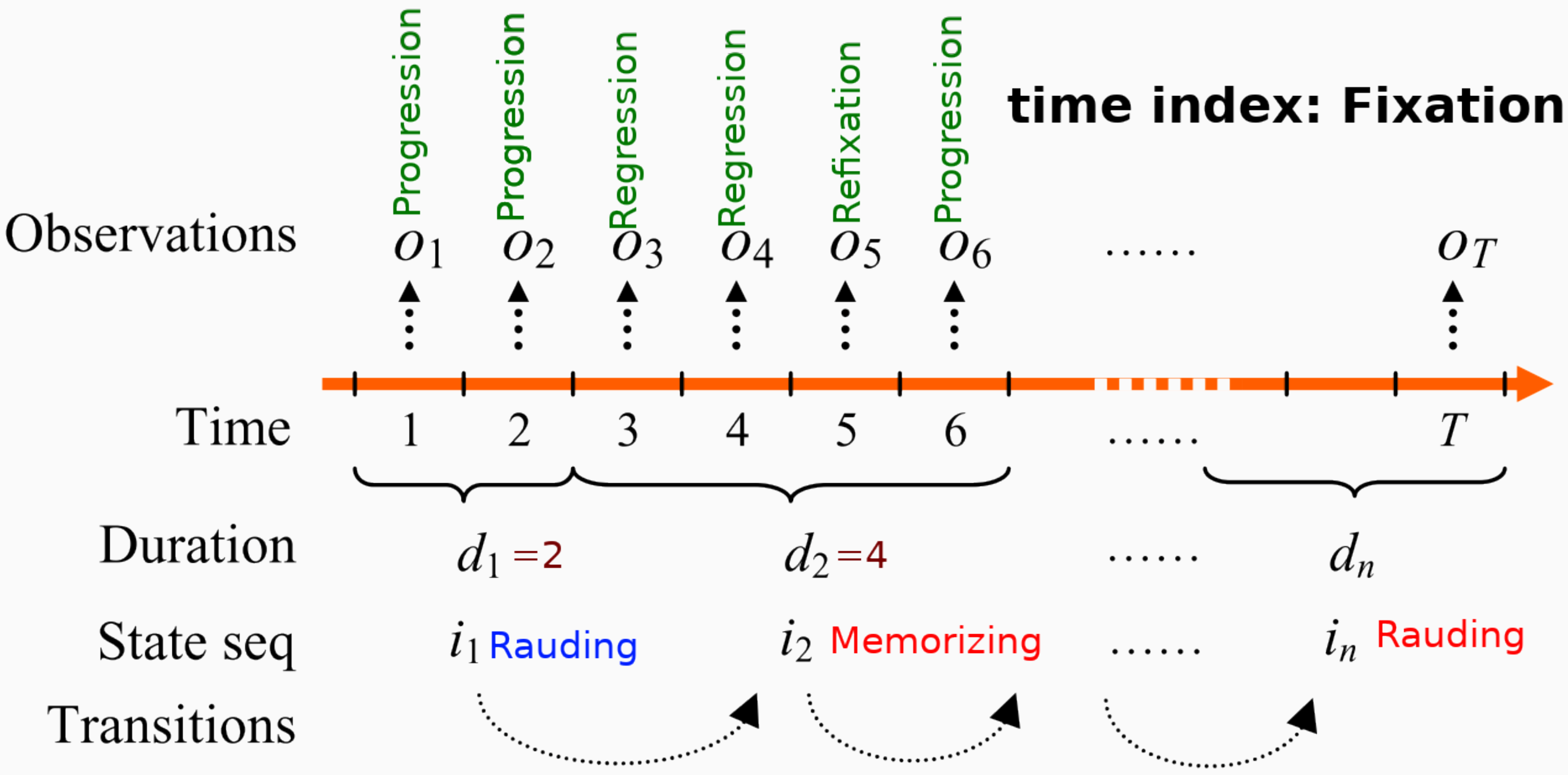
M-step: maximize with respect to θ

Restoration

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

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

Hidden semi-Markov Model (HSMM) (Yu, 2010)



$\mathcal{S} = \{Rauding, Skimming, Memorizing\}$

Duration sequence $\mathbf{D} = \{D_1, ..., D_T\}$

Sojourn distributions

$\forall d \in [\![1, \infty]\!]$

$p_j(d) =$

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

$\mathcal{V} = \{Regression, Refixation, Progression\}$

Transition probabilities

$\forall i \neq j \in \mathcal{S}$

$P(S_t = j | S_{t-1} = i, D_{t-1} = d) =$

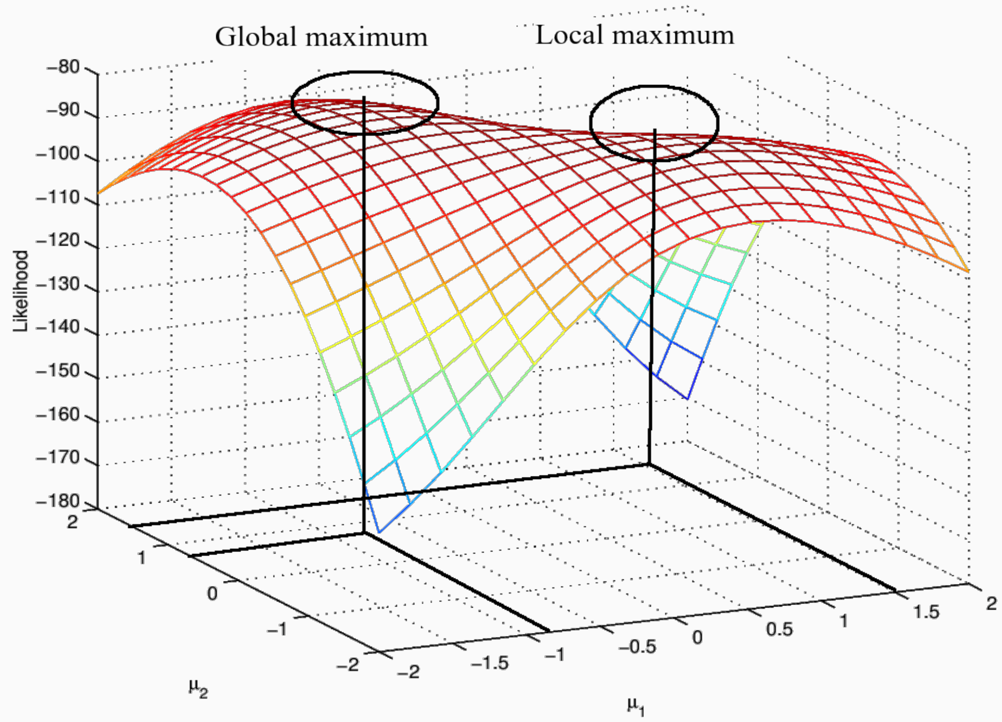
$$\begin{cases} A_{ij}, & \text{if } d = 1 \\ 0, & \text{if } d > 1 \end{cases}$$

Outline

Motivating example (Biernacki, Celeux, and Govaert, 2003)

- Consider 50 samples of a **Mixture Model** with two univariate Gaussian components, known proportions $\pi_1 = \pi_2 = 0.5$ and variances $\sigma_1^2 = 1, \sigma_2^2 = 1.5$ and unknown means $\mu_1 = -0.8, \mu_2 = 0.8$
- Maximum likelihood is performed with Expectation-Maximization algorithm
- Global maximum: 2 **distinct components** vs local maximum: 2 **similar components**

A two-mode likelihood surface (Biernacki, Celeux, and Govaert, 2003)



- EM is highly dependent on its starting position

Strategy 1: Perturbed Centers

- Let us consider **multiple observed sequences** $\mathbf{O} = \{\{O_t^{(1)}\}_{t\in[\![1,T_1]\!]}, \ldots, \{O_t^{(K)}\}_{t\in[\![1,T_K]\!]}\}$.

State of the art for HMMs:

- Juan, García-Hernández, and Vidal, 2004 showed experimentally that what seems to work best for BMMs is a **slightly perturbed version of the parameters** around their centers (**Equiprobability**). The hmmlearn python library uses the same technique for HMMs.

Strategy 1 for HSMMs: (control strategy)

- Initial parameters $\{a_{ij}, b_j(v_g), \pi_j\}$ were perturbed with a sample of **Dirichlet distribution** while $\{p_j(d)\}$ were set to Geometric distributions with parameter $p = 0.1$.

- **Issue:** how to perturb distributions with infinite support and unknown sojourn distribution ?

Strategy 2: Sequence Breaking Framework i

Algorithm 1: HighLikelihoodSearch: High local maximum of the likelihood search by sequence breaking

HighLikelihoodSearch(α, N)

Input: $\alpha \in [\![1, K]\!]$ the number of sequence to sample,
N, the number of initialization

for $n \leftarrow 1$ **to** N **do**
- Sample $\mathbf{O}^{(Q_\alpha)}$, α observed sequences from $\mathbf{O}$
- Generate $\mathbf{S}^{(Q_\alpha)}$, α corresponding state sequences with **SequenceBreaking**(α)
- Compute θ^{init}, the MLE for a semi-Markov chain on $(\mathbf{S}^{(Q_\alpha)}, \mathbf{O}^{(Q_\alpha)})$
- Compute $\hat{\theta}^{(n)}$ on $\mathbf{O}$ with EM using θ^{init} as a starting position

end

Choose $\hat{\theta}^*$, the set of parameters leading the highest likelihood

Output: $\hat{\theta}^*$, a high local maximum of the log-likelihood.

Strategy 2: Sequence Breaking Framework ii

Algorithm 2: SequenceBreaking

Input: α, a number of state sequence to generate

for $k \leftarrow 1$ **to** α **do**

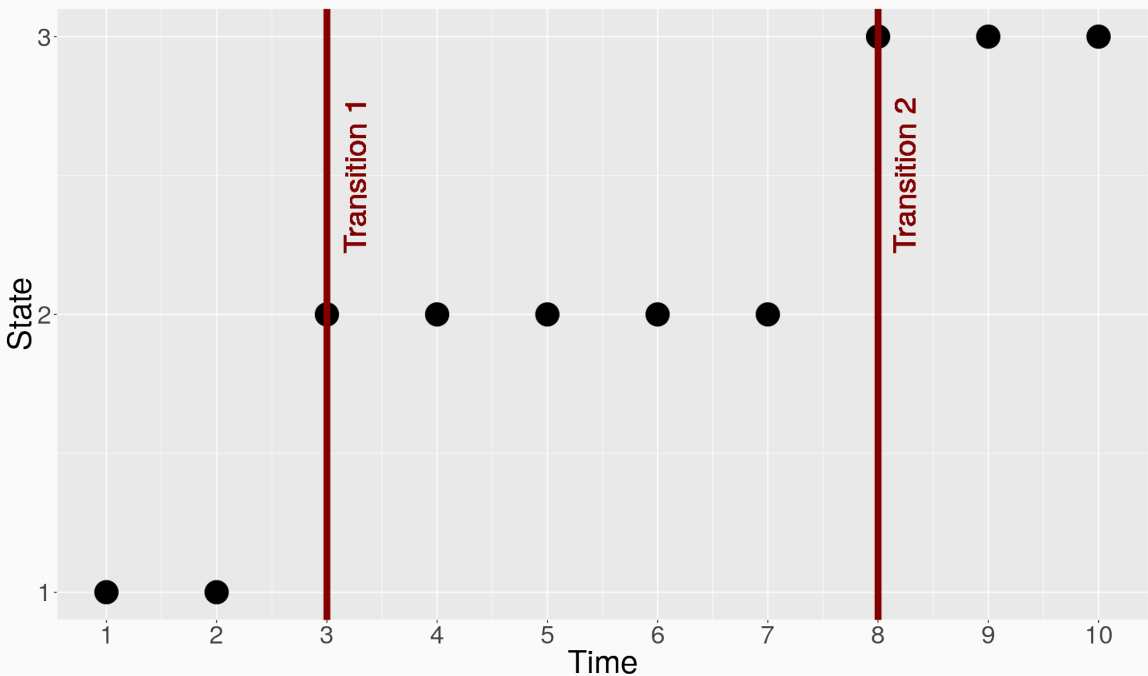
  Sample a number of transitions

  Sample the corresponding transition positions

  Fill the sequence $S^{(k)}$ by sampling states for each piece of sequence

end

Output: $S^{(Q_\alpha)}$, a randomly sampled state sequences



Example with 2 transitions

Datasets

1. **artificial:** 100 HSMM sequences of length 100, with $M = 5$ states and $G = 5$ factors of the observed process,
2. **artificial with noise:** 20% of the observations were replaced at random,
3. **eye-movement:** 2390 sequences of length 17 ± 8 with $G = 5$ and $M = 5$, with the output process taking values in $\mathcal{V} = \{$long regression, regression, refixation, progression, long pregression$\}$.

Results i

Global results: Means and standard deviations of maximum likelihood. $N = 100$ initializations, 1000 EM iterations, significance level 5%.

	artificial	noisy artificial	eye-movement
Sequence Breaking (SB)	$\mathbf{-15448 \pm 2.3}$	$\mathbf{-15785 \pm 4.5}$	-50567 ± 236
Perturbed Centers (PC)	-15452 ± 2.9	-15782 ± 2.6	-50592 ± 274

- Sequences Breaking (SB) performs significant better on artificial data but seems to be worse than Perturbed Center (PC) on eye-movement data

Local results: 100 initializations were split into 10 blocks of 10 and for each block the max were kept and compared.

	artificial	noisy artificial	eye-movement
SB > PC	**9/10**	**9/10**	6/10

- Sequence Breaking (SB) seems to perform better with a limited amount of initializations.

Results ii

Convergence speed: mean difference of EM iterations until convergence.

	artificial	noisy artificial	eye-movement
n_iter SB - n_iter PC	**-133**	**-305**	-44

- Sequence Breaking (SB) initial values are created using observations and happen to give **starting position with a higher likelihood**.

Advantages of Sequence Breaking

- This **strategy is generalizable** to continuous observed process and any kind of state sojourn distribution.
- **Randomness** of the initial values is **controlled** by the hyperparameter α, the number of sequences to sample.

Outline

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** $\in \mathcal{V} = \{< -1, -1, 0, 1, > 1\}$
- Corresponding empirical frequencies: $9\%, 2\%, 26\%, 22\%$ and 43%
- **Time index: fixation onset**

Latent Process

- **Reading strategies ”hidden”**, to be recovered through different patterns of readmode
- **Number of reading strategies unknown** and to be determined

Model covariates

- Fixation duration, Saccade amplitude, **Textual properties, EEGs**

Model selection - Choosing number of Classes

2 high likelihood search strategies (HLS) tested:

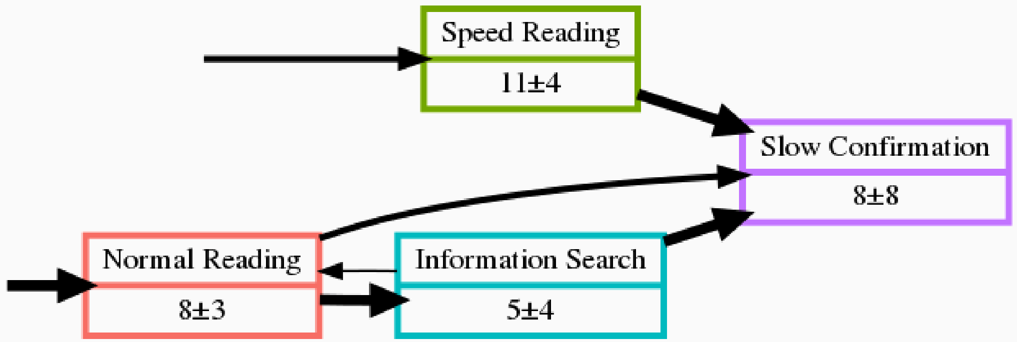
- Sequence breaking (SB),
- **Knowledge injection** (KI).

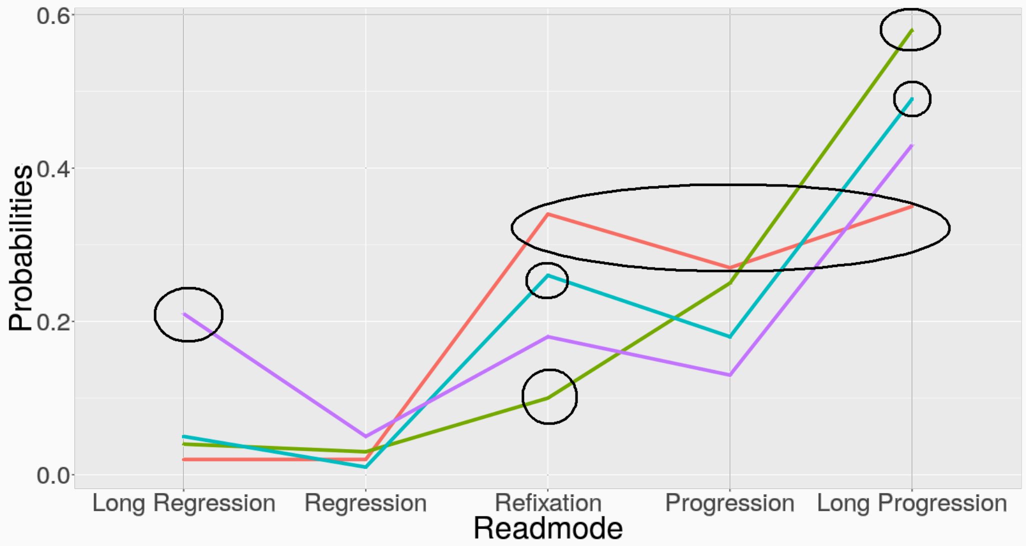
HLS strategy	#Classes	LL	BIC	Entropy	ICL*	#Params
KI	3	-50837	-101876	12426	-126728	19
KI	4	**-49745**	**-99798**	10771	-121341	29
KI	6	-50549	-101501	10067	-121635	38
SB	3	-50837	-101876	12426	-126728	19
SB	4	-50769	-101804	9758	-121321	26
SB	4	-50744	-101711	**5433**	**-112578**	21
SB	5	-50643	-101636	9255	-120146	33

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

- Moreover, a model **must be interpretable**

Estimated model parameters

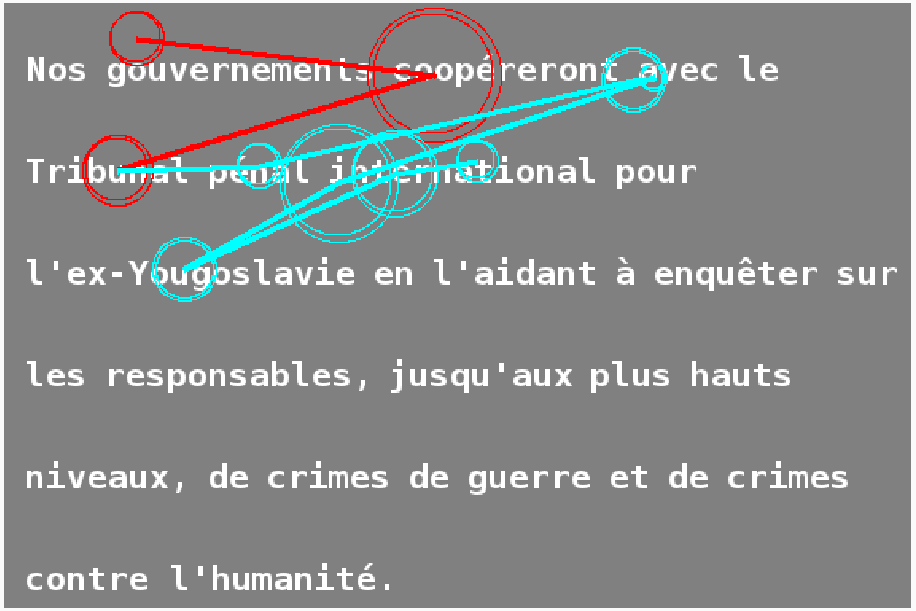




Each reading strategy is characterized by: a **readmode pattern**, a **sojourn distribution**, **probabilities to transit** to other reading strategies and an **initial probability**

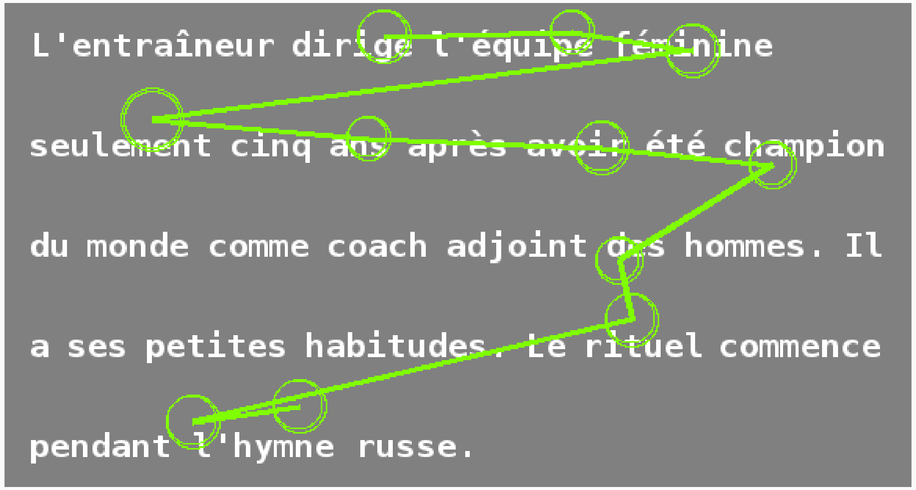
Scanpath restoration - HR / UR texts

"International tribunal" (Highly Related)



"The coach leads the women's team only five years after being world champion as assistant coach of men. He has his little habits. The ritual begins during the Russian anthem."

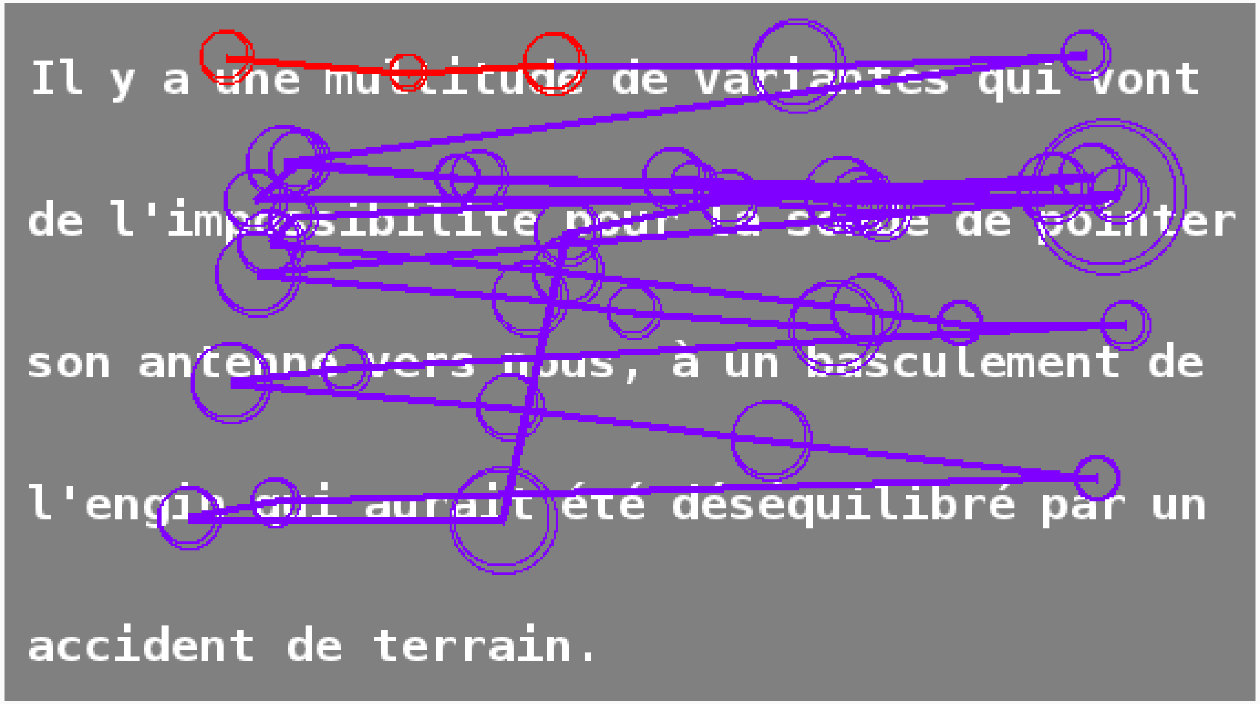
"Iraq conflict" (Unrelated related)



"The French football team defeated Australia in the World Cup final. The coach is very satisfied with his team and looks forward to future matches with enthusiasm and serenity."

Scanpath restoration - MR text

"Planets observation" (Moderately Related)

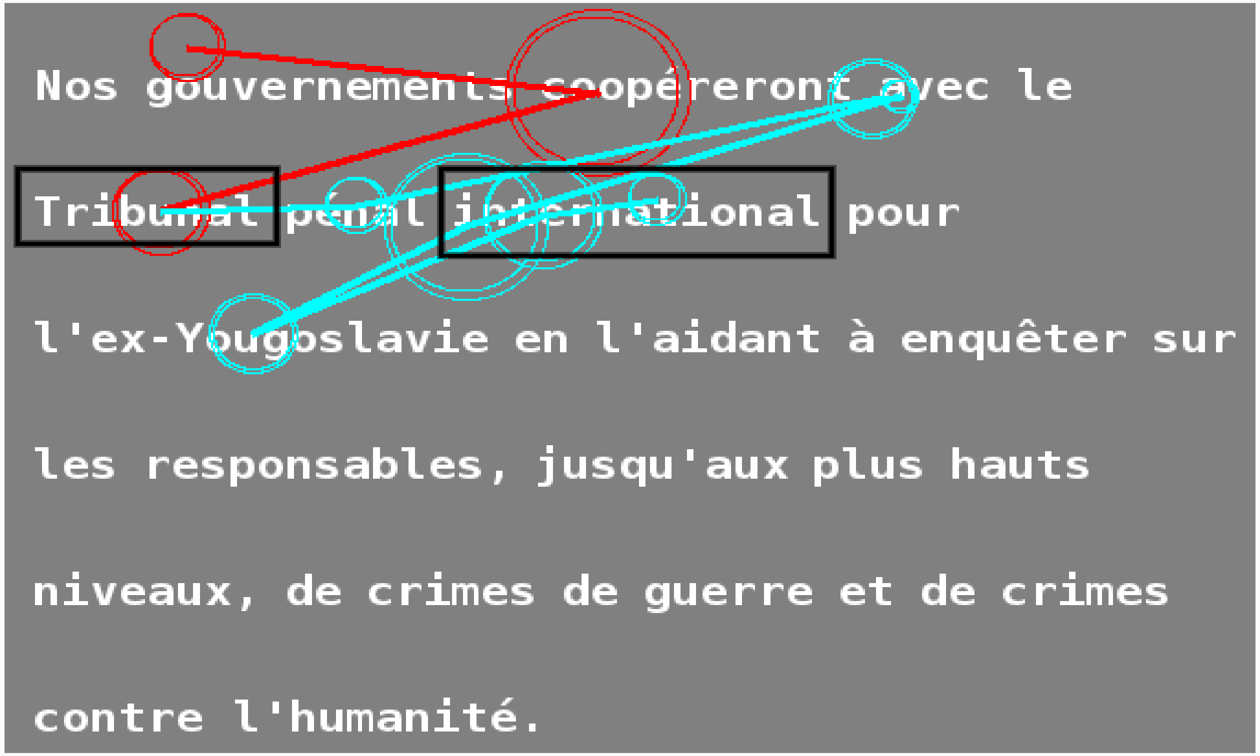


"There are a multitude of variants that range from the impossibility for the probe to point its antenna towards us, a tilting of the machine that would have been unbalanced by a terrain accident."

Outline

Text - Trigger word detection

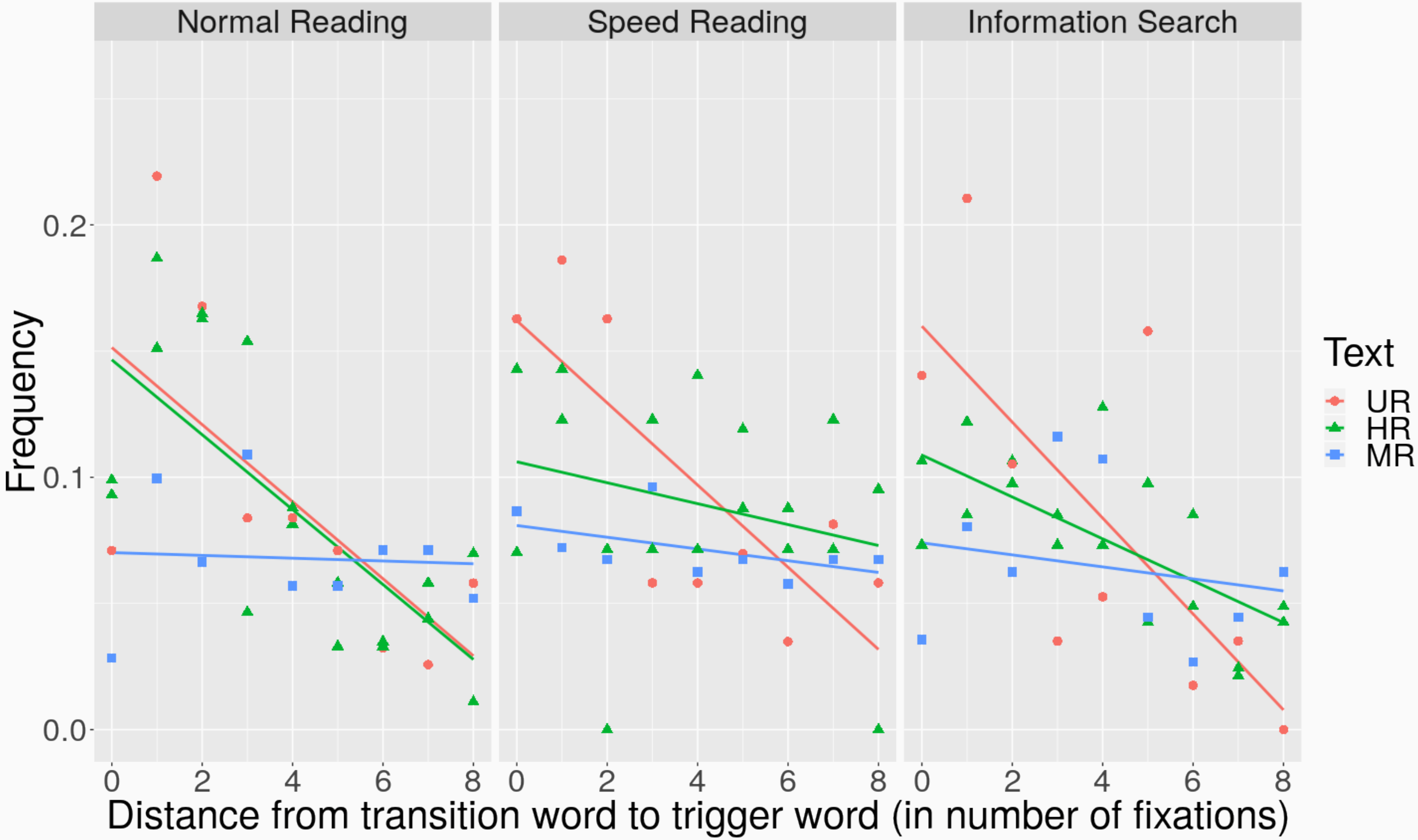
"International tribunal" (Highly related +)



"Our governments will collaborate with the International Criminal Tribunal for the former Yugoslavia in helping to investigate those responsible, to the highest level, for war crimes and crimes against humanity."

- **Do transitions occur around keywords ?**
- **Automatic detection of trigger words** w.r.t. topics
- We used Facebook's fastText word representations
- In HR texts, $w_{target} = \arg\max_i \frac{w_{topic} \cdot w_{text,i}}{\|w_{topic}\| \|w_{text,i}\|}$ where $w_.$ is a vector word representation and $w_{text,i}$ stands for i-th word in text.
- in UR texts, we added a log frequency factor for **specificity**

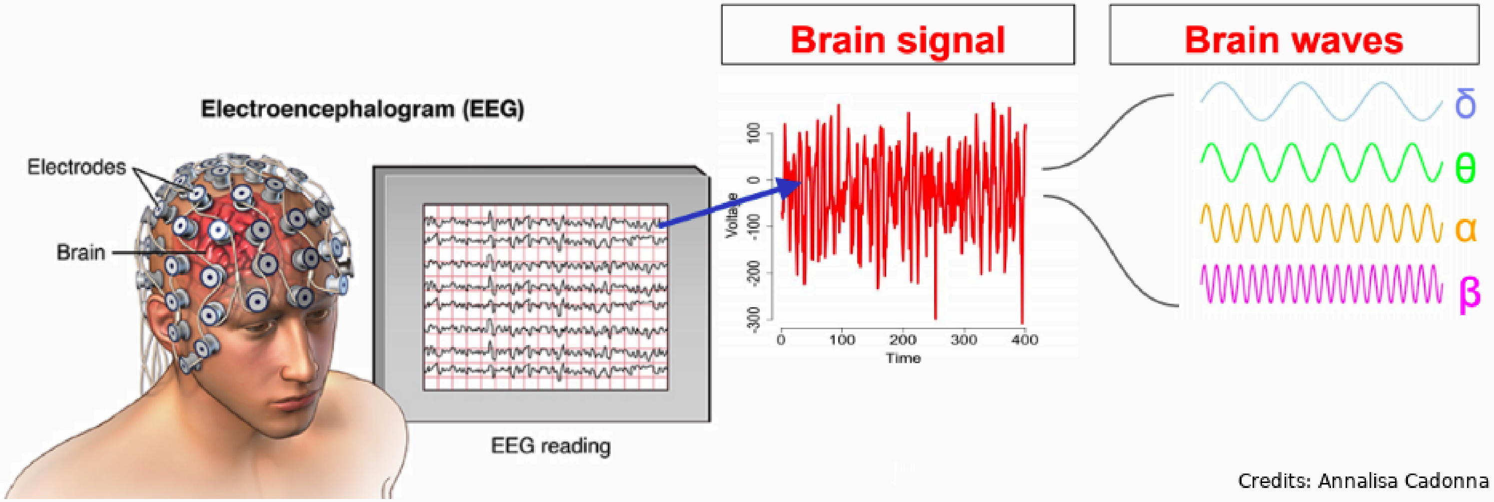
Text - Distance between target words to transitions



- Reading strategy **transitions occurs around keywords** when exiting states Normal Reading and Information search,
- the effect is less salient when exiting Speed Reading.

EEGs analysis - Bands, activities, tasks

- **Issue:** Eye fixation-related potential (EFRP) is difficult because of time-locking on uncertain and random reading strategies



Credits: Annalisa Cadonna

- **MODWT**: Decompose patterns that might not be visible on time domain
- α and θ oscillations reflect memory performances (Neuper and Klimesch, 2006),
- Left hemispheric lateralization for **verbal working memory** as well as right hemisphere lateralization for spatial working memory (Nagel et al., 2013)
- Memory encoding and restitution differences observed in α band (Seidkhani et al., 2017)

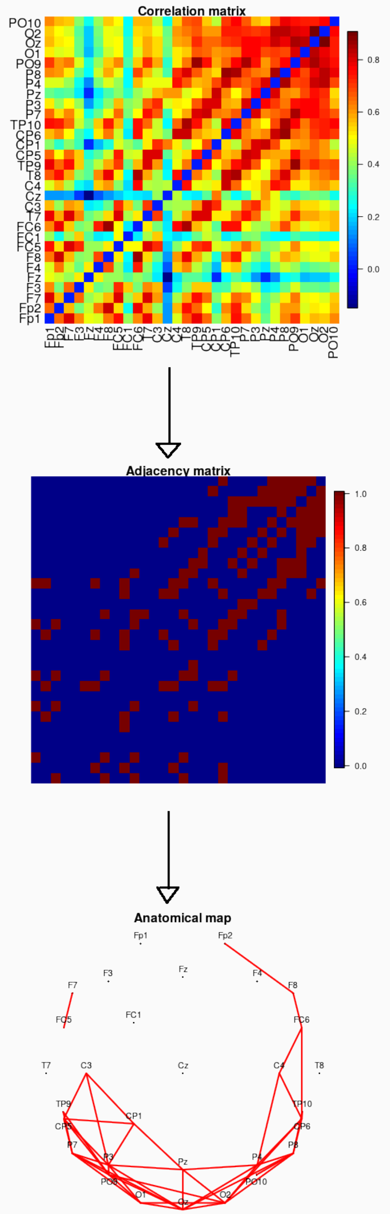
EEGs - Channel correlation for information diffusion analysis

Small-world network analysis (Achard et al., 2006)

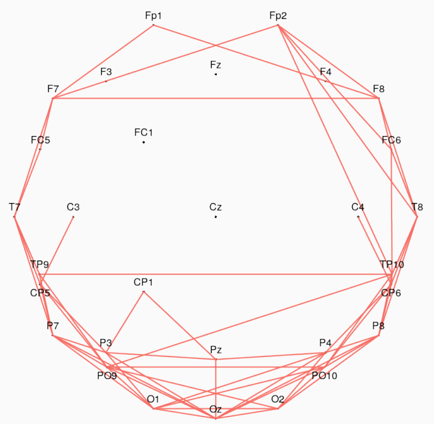
1. Maximal Overlap Discrete Wavelet Transform - LA8 filter
2. Wavelet correlation
3. Wavelet correlation confidence interval (Whitcher, Guttorp, and Percival, 2000) and **hypothesis testing**
4. Global **thresholding** into adjacency matrix & **Graph**
5. Graph metrics

Our reworked methodology:

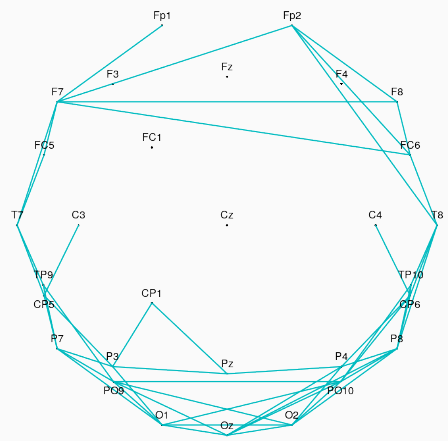
- **Wavelet correlations per phase** on the concatenation of the wavelet coefficients of all trials
- **Individual variability**: thresholding per individual to reduce the graph variability



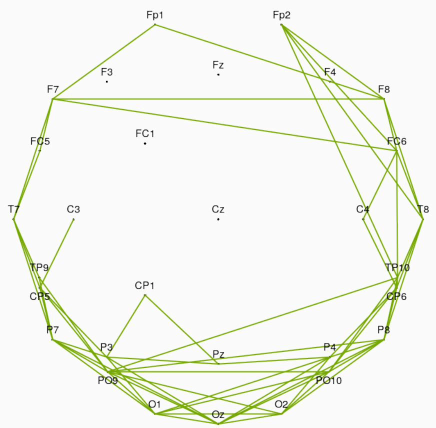
EEGs - Anatomical maps for most salient scale θ



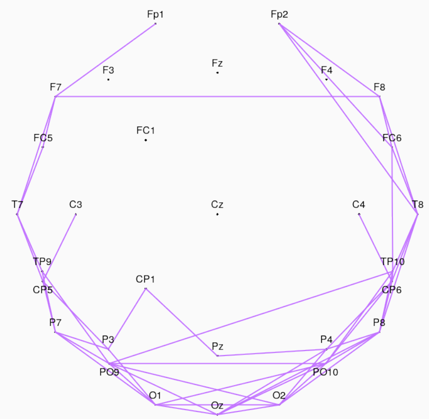
Normal reading - mean degree: 3.46, efficiency: 0.31



Information search - mean degree: 3.2, efficiency: 0.30



Speed reading - mean degree: 3.6, efficiency: 0.33

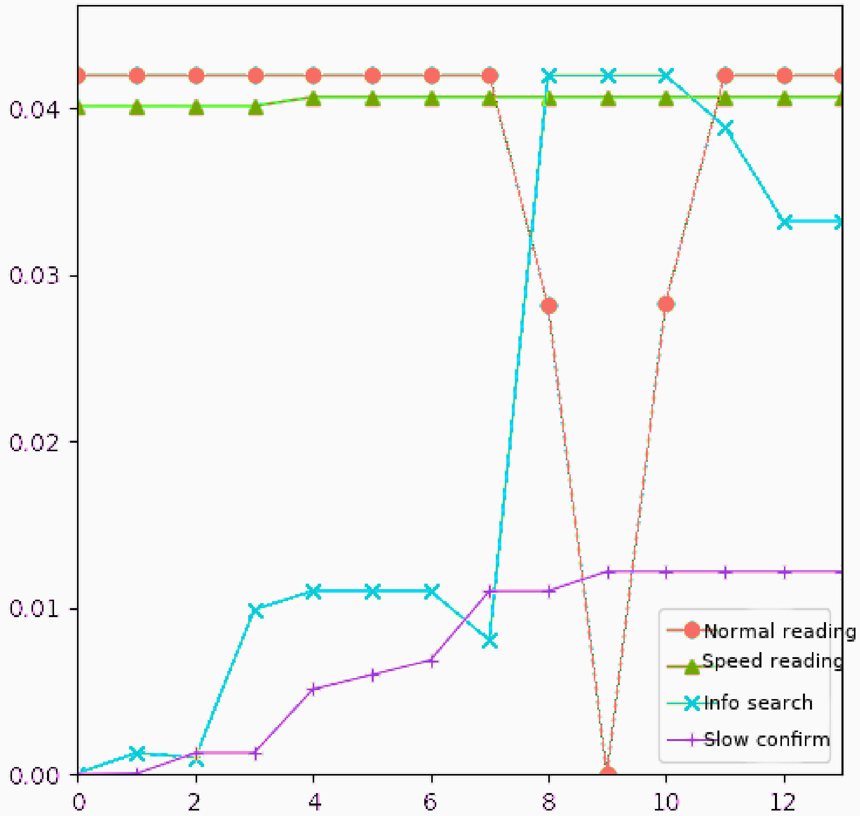


Slow confirmation - mean degree: 3, efficiency: 0.27

Issues - Uncertainty of state sequence restoration

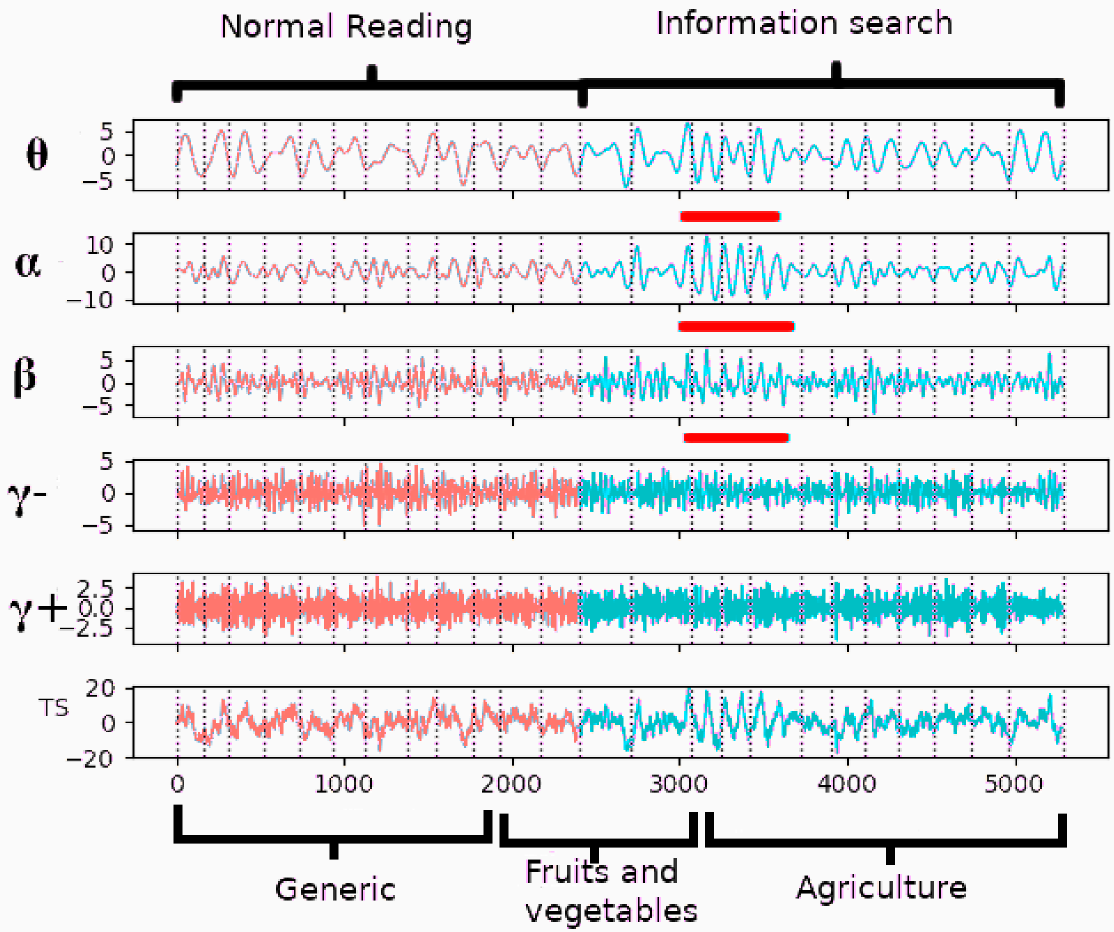
- State Entropy computation (Durand et al. 2012) has shown a high amount of uncertainty.
- Computation of Posterior probabilities of state sequence:

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



Issues - delay in EEGs

"Economic growth" - Unrelated text



- **Visible delay** between ocular and brain activities
- We want to use delayed EEGs to **reduce segmentation uncertainty**

Outline

Modeling specifications

- **Asynchronous Heterogeneous** Hidden semi-Markov Model (AHHSMM)

Different sampling rates

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

Delayed State

- Let $\{S_1^{(2)}, \dots, S_T^{(2)}\}$ a discrete latent state taking values in $\mathcal{S}$ and encoding the first SMC $\{S_1, \dots, S_{N_T}\}$ at a higher sampling rate, plus a lag.
- We denote the lag $\{\epsilon_{N_1}, \dots, \epsilon_{N_T}\}$, with $\epsilon_{N_t} \in \{1, \dots, \mathcal{L}\}$ in its most general form.
- Hence we have: $S_t^{(2)} = S_{N_t - \epsilon_{N_t}}, \forall t \in [\![\epsilon_1, \tau]\!]$.
- ϵ_{N_t} could be deterministic, random, autoregressive, conditional to channels or states. Its nature is to be determined.

AHHSM new parameters

Non-constant non-iid per state lag

- $v_{lk} = P(\epsilon_1 = l | S_1 = k), \quad \forall k \in \mathcal{S}, l \in \mathcal{L}$
- $\rho_{l'lk} = P(\epsilon_j = l' | \epsilon_{j-1} = l, S_j = k), \quad \forall k \in \mathcal{S}, l, l' \in \mathcal{L}$
  with j, the fixation index.

Output process related to EEGs models wavelet coefficients

- Let $O_t^{(2)}$ the concatenation of all wavelet coefficients and all channels at time t.
- $P(O_t^{(2)} | S_t^{(2)} = j) = \mathcal{N}(0, \Sigma_j)$, with Σ_j of size $(\#channels\#scales)^2$.
- We expect the covariance matrix to be **sparse**

Modified Expectation-Maximization algorithm

Let $\theta = (\theta_1, \theta_2)$, with $\theta_1 = \{\pi_k, A_{kk'}, p_k(d), b_k(v_g)\}$ and $\theta_2 = (\rho_{l'lk}, \mu_{lk}, \Sigma_j)$, the AHHSMM model parameters.

Decomposition of the Expectation

$$\begin{aligned}
&\mathbb{E}_{\Theta_1^{(m)}, \Theta_2^{(m)}}[\log P_{\theta_1,\theta_2}(\mathrm{O}, \mathrm{O}^{(2)}, \mathrm{S}, \mathrm{S}^{(2)}) | \mathrm{O}, \mathrm{O}^{(2)}] \\
&= \mathbb{E}_{\Theta_1^{(m)}, \Theta_2^{(m)}}[\log P_{\theta_1}(\mathrm{O}, \mathrm{S}) + \log P_{\theta_2}(\mathrm{O}^{(2)} | \mathrm{S}^{(2)}) + \log P_{\theta_2}(\mathrm{S}^{(2)} | \mathrm{S}) | \mathrm{O}, \mathrm{O}^{(2)}],
\end{aligned}$$

Proposition

Let $(\Theta_1^{(m)}, \Theta_2^{(m)})_{m \geqslant 1}$ denote the sequence of iterates of the following modified EM algorithm and $(\tilde{\theta}_1, \tilde{\theta}_2)$ denote the true parameters.

$$\begin{aligned}
(\Theta_1^{(m+1)}, \Theta_2^{(m+1)}) = \Bigg(&\arg\max_{\theta_1} \mathbb{E}_{\Theta_1^{(m)}}[\log P_{\theta_1}(\mathrm{O}, \mathrm{S}) | \mathrm{O}], \\
&\arg\max_{\theta_1, \theta_2} \mathbb{E}_{\Theta_1^{(m)}, \Theta_2^{(m)}}[\log P_{\theta_2}(\mathrm{O}^{(2)} | \mathrm{S}^{(2)}) + \log P_{\theta_2}(\mathrm{S}^{(2)} | \mathrm{S}) | \mathrm{O}, \mathrm{O}^{(2)}] \Bigg),
\end{aligned}$$

Then $\lim_{m \to \infty} (\Theta_1^{(m)}, \Theta_2^{(m)}) = (\tilde{\theta}_1, \tilde{\theta}_2)$, *under the assumption that the MLE is consistent.*

Inference differences wrt HSMM

In E-step:

- A **different inference algorithm** for each lag hypothesis,
- the expected sufficient statistics (ESS) related to HSMMs remain the same,
- with the non-constant, non-iid per state lag, **3 new ESS** are introduced,
- one **extra hypothesis** is required: at each time step, the delay must be upper-bounded by the current fixation duration,
- **Efficient inference algorithm** in $\mathbf{O(T\mathcal{L}M^2\mathcal{D})}$ vs $O(TM^2\mathcal{D})$ for HSMM vs $O(TM^2)$ for HMM.

Outline

Contributions

- We proposed and assessed a promising initialization strategies for HSMMs' EM,
- 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 into a single probabilistic model.

Perspectives

- **Assessing initialization strategies for EM:** compare the strategies on a wide variety of datasets with different types of signals.
- **Individual variability:** Assess and correct individual variability of EGGs with mixed effect models.
- **EEGs**: strenghten results with more litterature.
- **Model testing:** Get preliminary results on simulations for the proposed model to ensure identifiability and evaluate the proposed modified EM procedure.
- **Scaling:** new model highly complex + high usage of RAM ($\sim$45Gb) to store MODWT coefficients of EEGs –> Online or preferably minibatch version.

Thank you

References i

Achard, Sophie et al. (2006). “A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs”. In: *Journal of Neuroscience* 26.1, pp. 63–72.

Biernacki, Christophe, Gilles Celeux, and Gérard Govaert (2000). “Assessing a mixture model for clustering with the integrated completed likelihood”. In: *IEEE transactions on pattern analysis and machine intelligence* 22.7, pp. 719–725.

– (2003). “Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models”. In: *Computational Statistics & Data Analysis* 41.3-4, pp. 561–575.

Carver, Ronald P (1990). *Reading rate: A review of research and theory.* Academic Press.

Frey, Aline et al. (2013). “Decision-making in information seeking on texts: an eye-fixation-related potentials investigation”. In: *Frontiers in systems neuroscience* 7, p. 39.

Juan, Alfons, José García-Hernández, and Enrique Vidal (2004). "EM initialisation for Bernoulli mixture learning". In: *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, pp. 635–643.

Nagel, Bonnie J et al. (2013). "Hemispheric lateralization of verbal and spatial working memory during adolescence". In: *Brain and cognition* 82.1, pp. 58–68.

Neuper, Christa and Wolfgang Klimesch (2006). *Event-related dynamics of brain oscillations*. Vol. 159. Elsevier.

Rayner, Keith (1998). "Eye movements in reading and information processing: 20 years of research.". In: *Psychological bulletin* 124.3, p. 372.

Reichle, Erik D et al. (1998). "Toward a model of eye movement control in reading.". In: *Psychological review* 105.1, p. 125.

Seidkhani, Hossein et al. (2017). "Task modulates functional connectivity networks in free viewing behavior". In: *NeuroImage* 159, pp. 289–301.

Simola, Jaana, Jarkko Salojärvi, and Ilpo Kojo (2008). "Using hidden Markov model to uncover processing states from eye movements in information search tasks". In: *Cognitive systems research* 9.4, pp. 237–251.

Volant, Stevenn et al. (2012). "Hidden Markov Models with mixtures as emission distributions". In: *arXiv preprint arXiv:1206.5102*.

Whitcher, Brandon, Peter Guttorp, and Donald B Percival (2000). "Wavelet analysis of covariance with application to atmospheric time series". In: *Journal of Geophysical Research: Atmospheres* 105.D11, pp. 14941–14962.

Yu, Shun-Zheng (2010). "Hidden semi-Markov models". In: *Artificial intelligence* 174.2, pp. 215–243.