

Indexer la complexité linguistique en lien avec la dynamique cérébrale à partir de signaux MEG

Soutenance de stage de fin d'étude d'ingénieur

Lucas Becquet

Ecole Centrale de Marseille

2023

Contexte

Laboratoire de Neuroscience Cognitive de Marseille (LNC)



Contexte

L'équipe Développement Informatique et Infrastructure Système pour les Sciences du Cerveau (DI²S²C)

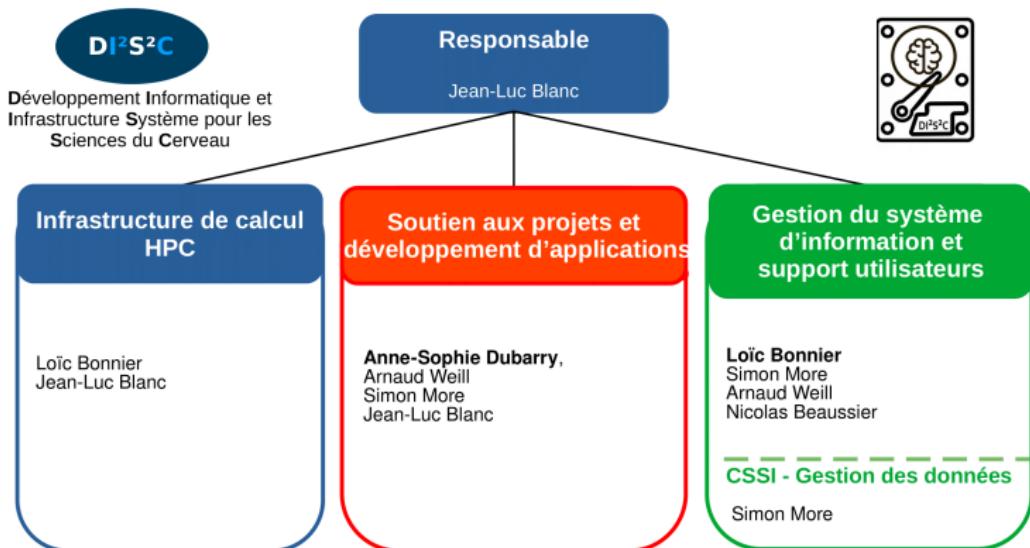


Figure – Organigramme de l'équipe DISC

Peut-on discriminer des conditions expérimentales et rendre compte d'une compréhension linguistique sur la base de la quantification précise de la dynamique cérébrale ?

Sommaire

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques

La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques

La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques

La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

De l'activité neuronale aux signaux neurophysiologiques

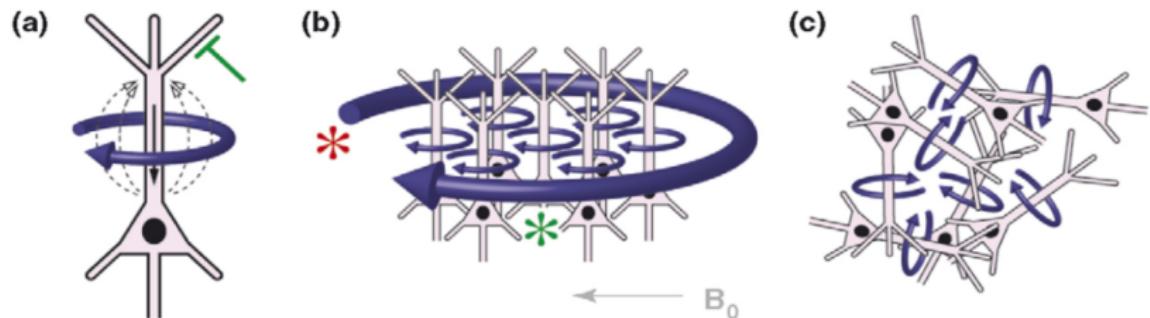


Figure – Activation neuronale. (a) Neurone seul avec représentation de son potentiel électrique (ligne pointillée) et du champ magnétique induit (flèche bleue). (b) Groupe de neurones orientés dans le même sens, leur activation synchrone crée un champ magnétique résultant. (c) Neurones orientés de manière aléatoire, les champs magnétiques s'annulent. Adapté de Jasanof, 2007

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques

La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

La technique d'enregistrement MEG

David Cohen et James Zimmerman, 1960

Champs magnétiques de 10^{-15} Tesla

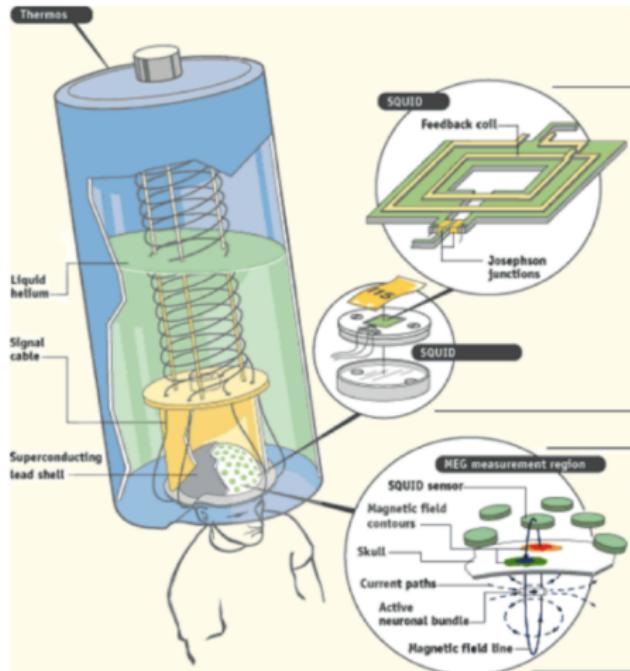


Figure – Illustration de la MEG. Schéma d'un système MEG complet et de ses composants (issu de Dubarry, 2016)

La technique d'enregistrement MEG

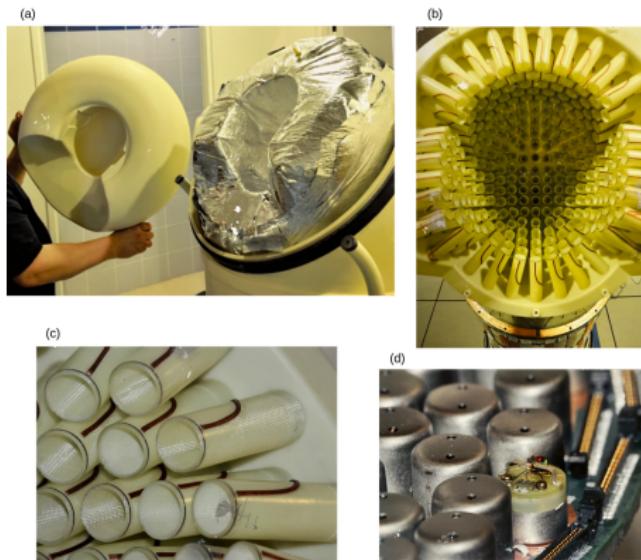


Figure – Photographies des différents composants du système d'acquisition MEG, prises à différentes profondeurs. Centre MEG de la Timone, Marseille

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques

La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Description du dataset

A 204-subject multimodal neuroimaging dataset to study language processing, Jan-Mathijs Schoffelen and Robert Oostenveld and Nietzsche H.L.Lam and Julia Uddén and Annika Hultén and Peter Hagoort, 2019.



N=204



Resting state measurement
Language task
Diffusion weighted imaging
Structural imaging



Resting state measurement
Language task



N=102



N=102



Figure – Organisation du dataset

Description du dataset

1. 180 phrases simples appelées simple relative clause sentence (RC-)
2. 180 phrases complexes appelées relative clause sentence (RC+)

	Sentence	Word list
Complex (Relative Clause, RC+)	Het aardige vrouwtje gaf Henk die een kleurige papegaai gekocht had een zak pitjes <i>The nice lady gave Henk, who had bought a colorful parrot, a bag seeds.</i>	Zak een kleurige aardige een had die vrouwtje papegaai gaf het gekocht pitjes Henk <i>Bag a colorful nice a had who lady parrot gave the bought seeds Henk</i>
Simple (RC-)	Dit zijn geen regionale problemen zoals die op de Antillen. <i>These are no regional problems such as those on the Antilles.</i>	zoals geen die Antillen problemen regionale zijn de dit op <i>such as no those Antilles problems regional are the these on</i>

Figure – Exemple des différents stimuli : Phrase simple (RC- sentence), phrase complexe (RC+ sentence) et listes aléatoires de mots associées

Description du dataset

Trigger value	meaning
1	Onset of individual word (visual task) or first word (auditory task) in a Relative Clause containing sentence (RC+).
2	Onset of 'target' word in a RC+ sentence.
3	Onset of individual word (visual task) or first word (auditory task) in a word list derived from a RC+ sentence.
4	Onset of 'target' word in a word list derived from a RC+ sentence.
5	Onset of individual word (visual task) or first word (auditory task) in a sentence without a relative clause (RC-).
6	Onset of 'target' word in a RC- sentence.
7	Onset of individual word (visual task) or first word (auditory task) in a word list derived from a RC- sentence.
8	Onset of 'target' word in a word list derived from a RC- sentence.
10	Mini block instruction stimulus 'WOORDEN' (words) or 'ZINNEN' (sentences)
11	Response (index), auditory task (in visual task, this event has value 1)
12	Response (middle), auditory task (in visual task, this event has value 2)
13	Experimenter response to continue after break, auditory task (in visual task, this event has value 3)
14	Start of audio file (auditory task only)
15	Offset of word picture (visual task) or audio file (auditory task)
20	Fixation picture, pre-trial baseline.
30	Pause
40	Question

Figure – Identification des différents évènements et stimuli lors de l'enregistrement de la tâche de compréhension

Description du dataset

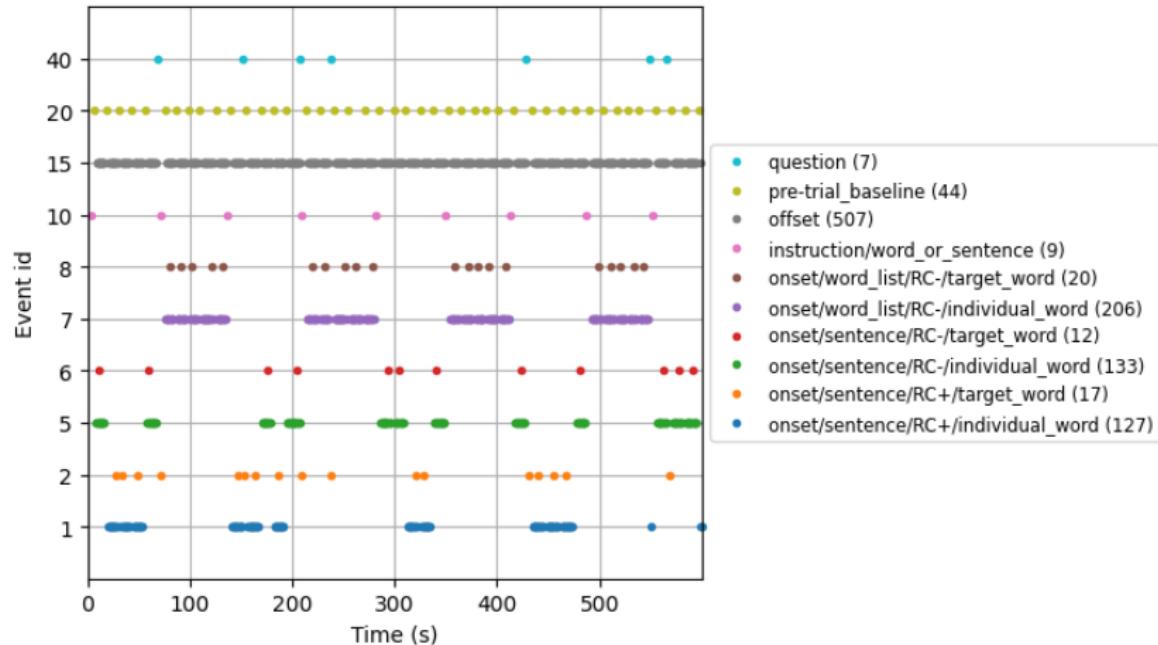


Figure – Visualisation des différents événements au cours du temps d'un sujet de la tache visuelle

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques

La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Pré-traitement des signaux MEG

MNE-Python

Measurement date	January 01, 1970 10:46:00 GMT
Experimenter	Unknown
Participant	Anonymized297739_1534851269_0
Digitized points	7 points
Good channels	3 Stimulus, 44 misc, 28 Reference Magnetometers, 273 Magnetometers, 4 EEG
Bad channels	None
EOG channels	Not available
ECG channels	Not available
Sampling frequency	1200.00 Hz
Highpass	0.00 Hz
Lowpass	600.00 Hz
Filenames	sub-V1001_task-visual_meg.meg4 sub-V1001_task-visual_meg.1_meg4 sub-V1001_task-visual_meg.2_meg4 sub-V1001_task-visual_meg.3_meg4
Duration	01:07:08 (HH:MM:SS)

Figure – Meta-données d'un enregistrement de la tâche de compréhension visuelle. On peut observer les différents canaux, y accéder aussi plus en détails, on voit aussi que la fréquence d'échantillonnage est de 1200Hz

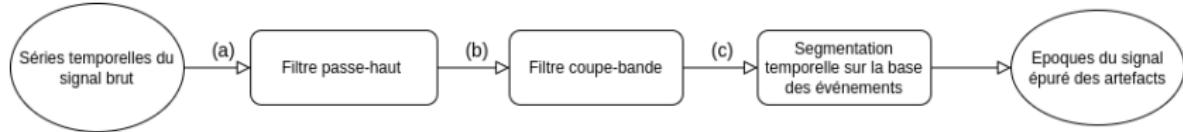


Figure – Protocole de pré-traitement des données MEG. (a) On applique d'abord un filtre passe-haut. (b) On applique un filtre coupe-bande. (c) On effectue la segmentation temporelle et donc la création d'époques des séries temporelles du signal MEG

Notions de neurophysiologie

- De l'activité neuronale aux signaux neurophysiologiques
- La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

- Séries temporelles et filtre passe-haut
- Spectre et filtre coupe-bande
- Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

- Représentation symbolique d'un système dynamique
- Entropie
- Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

- Target word vs individual word
- Phrases simples vs listes aléatoires de mots issues de phrases simples
- Tâche visuelle vs tâche auditive

Séries temporelles et filtre passe-haut

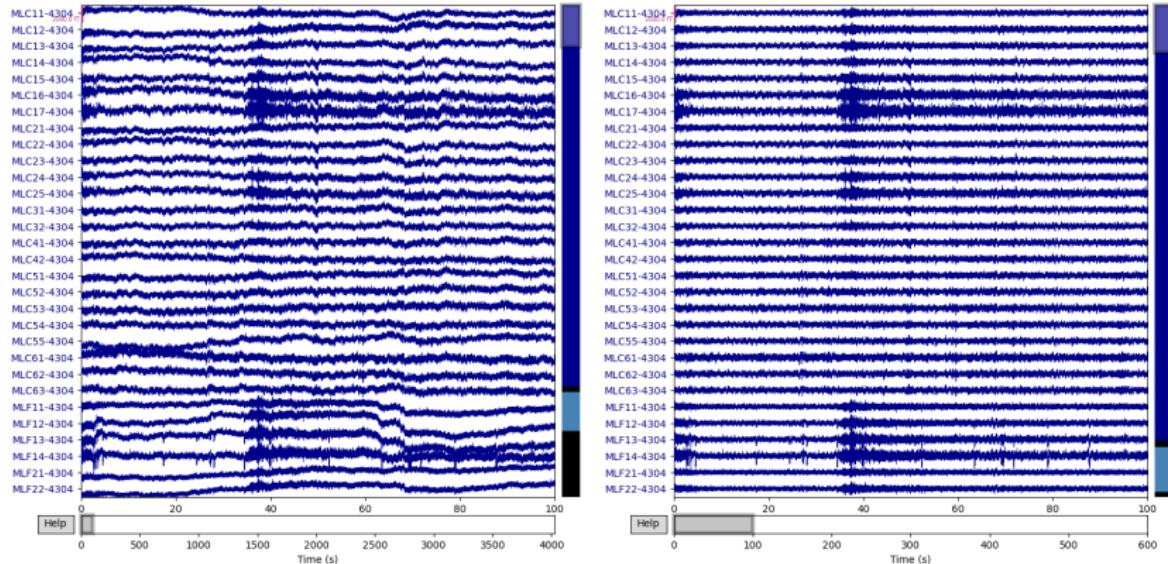


Figure – A gauche, séries temporelles des différents canaux du signal MEG brut. A droite, séries temporelles après application du filtre passe-haut

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques
La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Spectre et filtre coupe-bande

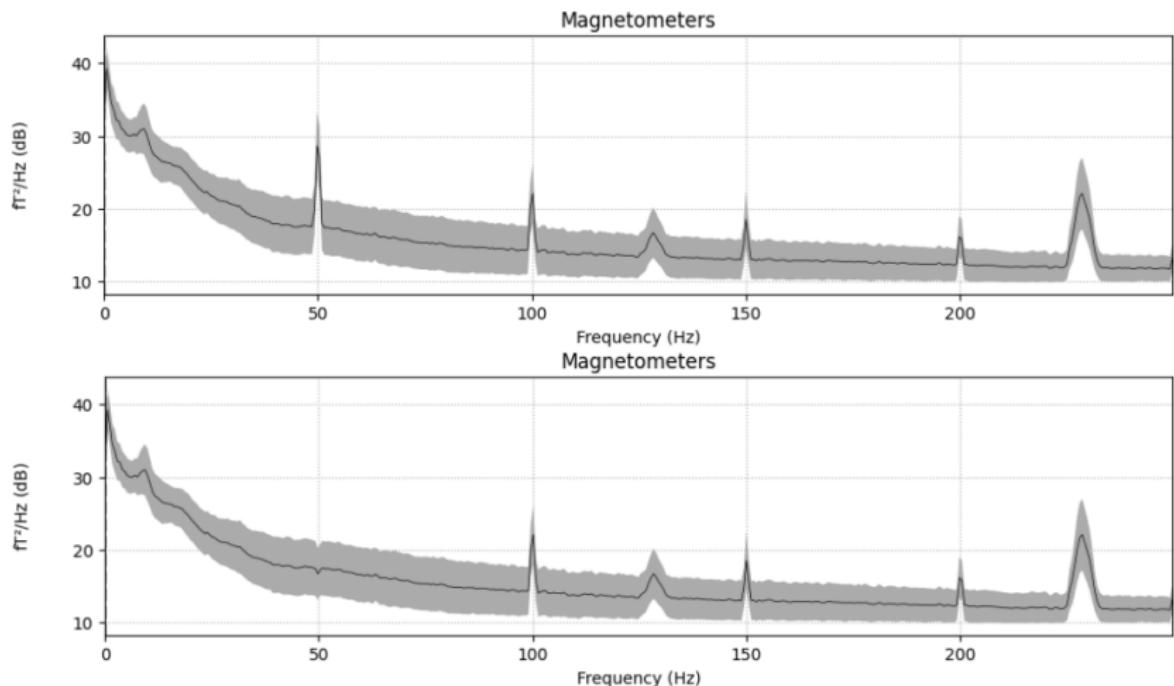


Figure – Graphe de la densité spectrale relative à l'enregistrement d'un sujet de la tâche visuelle. En haut, le spectre du signal brut. En bas, le spectre du signal une fois le filtre coupe-bande à 50 Hz appliqué

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques
La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

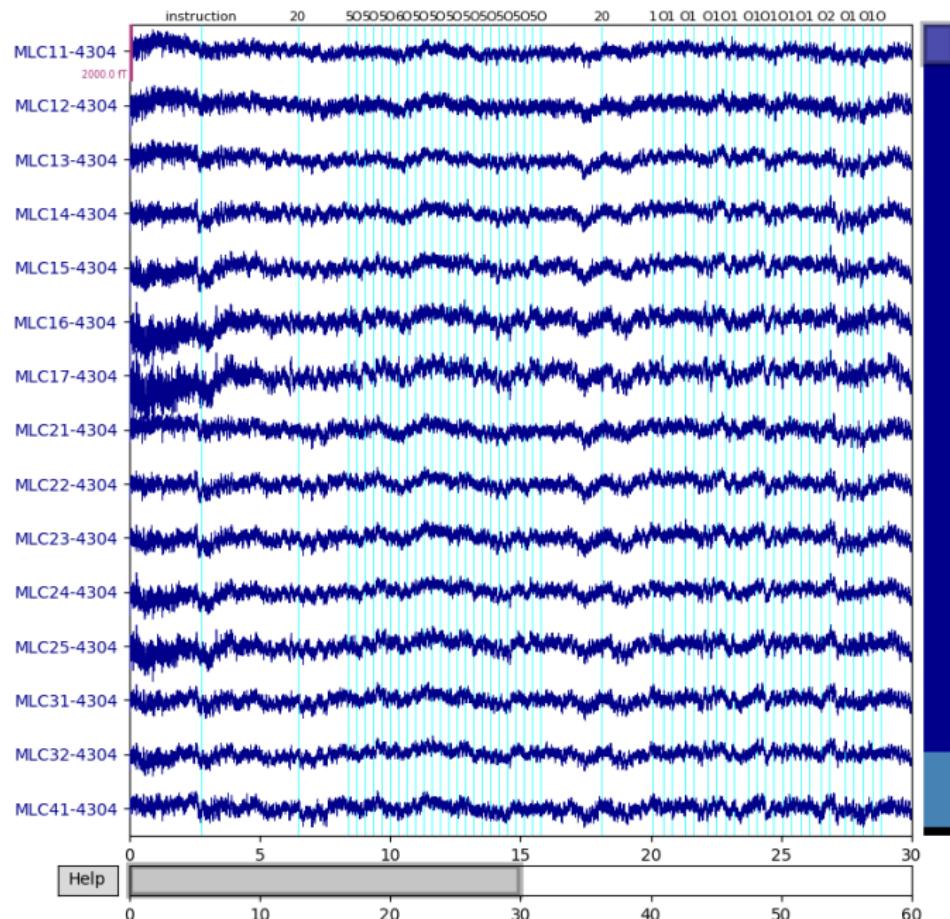
Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Segmentation temporelle du signal



Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques

La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Schéma de l'algorithme mis en place durant le stage

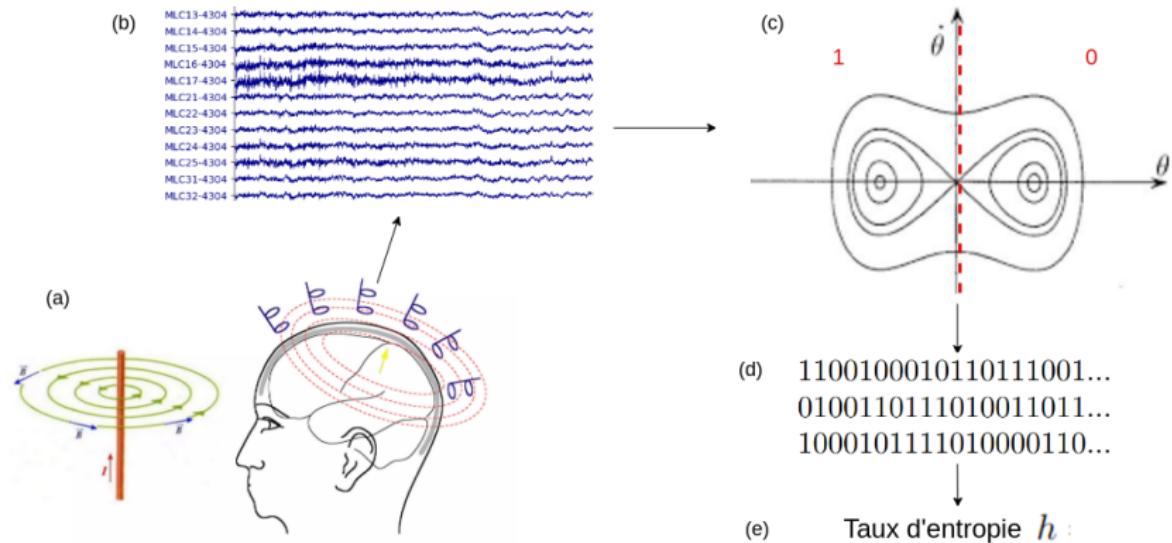


Figure – Schéma de l'algorithme mis en place durant le stage

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques
La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut
Spectre et filtre coupe-bande
Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique
Entropie
Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word
Phrases simples vs listes aléatoires de mots issues de phrases simples
Tâche visuelle vs tâche auditive

Représentation symbolique d'un système dynamique - Théorie

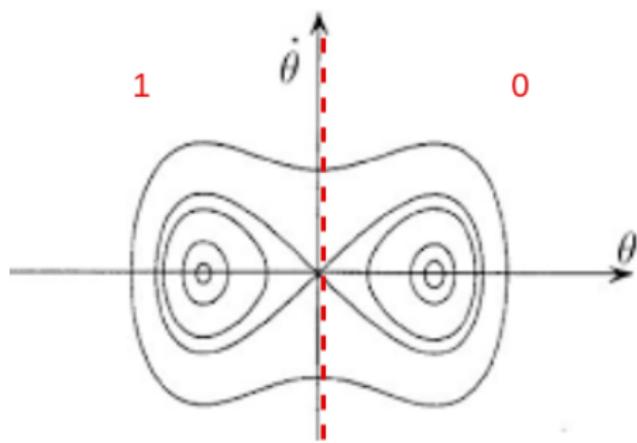


Figure – Illustration de la représentation symbolique dans l'espace des phases. Ici représenté en seulement 2 dimensions avec une partition binaire, les points des trajectoires à gauche de la ligne en pointillé rouge sont associés au symbole 1 tandis que ceux à droite sont associés au symbole 0.

Représentation symbolique d'un système dynamique - Pratique

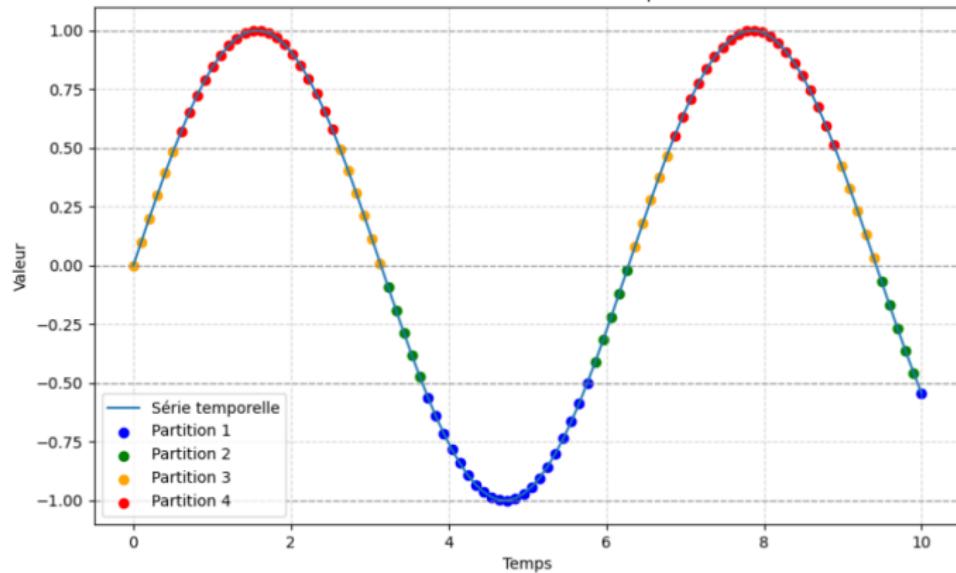


Figure – Illustration d'une équipartition par histogramme des valeurs d'une série temporelle. Exemple d'une partition à 4 symboles d'une sinusoïde. On peut identifier les différents intervalles qui correspondent chacuns à un symbole de la partition

Représentation symbolique d'un système dynamique - Pratique

- ▶ Matrice des données (capteurs × temps)

Représentation symbolique d'un système dynamique - Pratique

- ▶ Soit M la transposée de la matrice des données $m \times n$,

Représentation symbolique d'un système dynamique - Pratique

- ▶ Soit M la transposée de la matrice des données $m \times n$,
- ▶

$$M = USV^T \quad (1)$$

Représentation symbolique d'un système dynamique - Pratique

- ▶ Soit M la transposée de la matrice des données $m \times n$,



$$M = USV^T \quad (1)$$

- ▶
 1. $U \in \mathbb{R}^{m \times n}$
 2. S est une matrice diagonale de $\mathbb{R}^{n \times n}$ qui contient les valeurs propres de $M^T M$
 3. $V^T \in \mathbb{R}^{n \times n}$

Représentation symbolique d'un système dynamique - Pratique

Critère de sélection de Schwarz

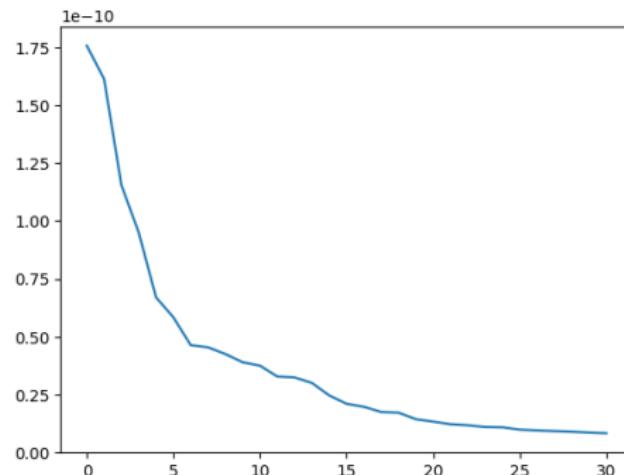
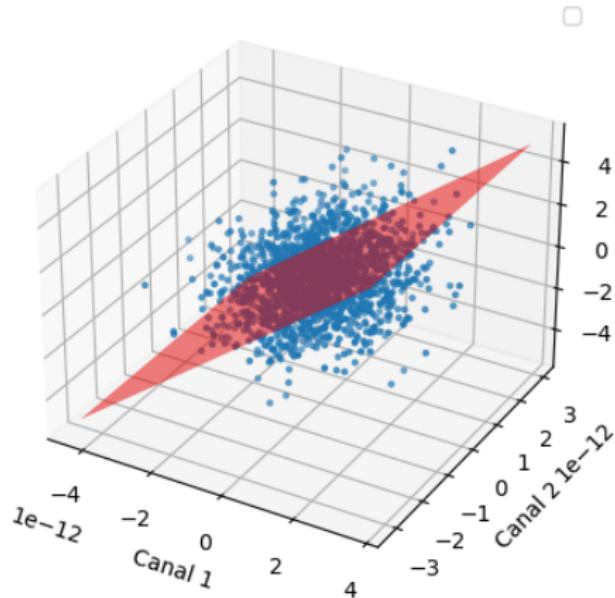


Figure – Graphe des 30 premières valeurs propres issues de la SVD de la matrice des données pour un sujet de la tâche visuelle

Représentation symbolique d'un système dynamique - Pratique

On ne garde donc qu'une partie d des vecteurs propres contenus dans V de sorte à avoir une matrice $U' \in R^{n \times d}$. De la même manière, on crée $S' \in R^{d \times n}$ en tronquant S et $V' \in R^{n \times d}$. On obtient alors notre matrice des données réduite par $M' = U'S'V'^T$.



Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques

La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Entropie

Soit X une variable aléatoire discrète ayant \mathcal{X} comme alphabet.

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x) \quad (2)$$

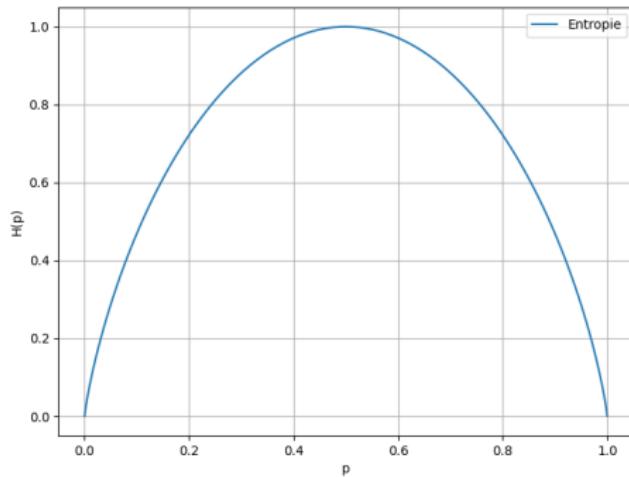


Figure – Entropie $H(p)$ en fonction de la probabilité p d'un événement binaire

Entropie

Théorie de l'information

Quantité d'information produite par la source

Théorie des systèmes dynamiques

Complexité intrinsèque du système dynamique

Informatique théorique

Longueur minimale d'un code universel pour transmettre un message

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques
La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut
Spectre et filtre coupe-bande
Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique
Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word
Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Définition de l'entropie dans notre contexte

$$H_n = - \sum_w p_n(w) \ln p_n(w) \quad (3)$$

1. w mot de longueur n
2. $p_n(w)$ distribution de probabilité

Taux d'entropie

$$h = \lim_{n \rightarrow \infty} \frac{H_n}{n} = \lim_{n \rightarrow \infty} H_{n+1} - H_n \quad (4)$$

Estimateur algorithmique de Lempel-Ziv



$1 \cdot 0 \cdot 01 \cdot 11 \cdot 100 \cdot 101 \cdot 00 \cdot 010 \cdot 11 \dots$ (5)

Estimateur algorithmique de Lempel-Ziv



$$1 \cdot 0 \cdot 01 \cdot 11 \cdot 100 \cdot 101 \cdot 00 \cdot 010 \cdot 11 \dots \quad (5)$$



$$\hat{L} = \frac{\mathcal{N}_w[1 + \log_k \mathcal{N}_w]}{N} \quad (6)$$

Estimateur algorithme de Lempel-Ziv



$$1 \cdot 0 \cdot 01 \cdot 11 \cdot 100 \cdot 101 \cdot 00 \cdot 010 \cdot 11 \dots \quad (5)$$



$$\hat{L} = \frac{\mathcal{N}_w[1 + \log_k \mathcal{N}_w]}{N} \quad (6)$$

► où

$$\lim_{n \rightarrow \infty} \hat{L} = \frac{h}{\ln k} \quad (7)$$

1. k taille de l'alphabet
2. \mathcal{N}_w nombre de mots du dictionnaire
3. N taille de la séquence symbolique

Notions de neurophysiologie

- De l'activité neuronale aux signaux neurophysiologiques
- La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

- Séries temporelles et filtre passe-haut
- Spectre et filtre coupe-bande
- Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

- Représentation symbolique d'un système dynamique
- Entropie
- Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

- Target word vs individual word
- Phrases simples vs listes aléatoires de mots issues de phrases simples
- Tâche visuelle vs tâche auditive

Scikits-symbolic

Module python : fonctions, de classes et de méthodes de la théorie de l'information permettant de manipuler les séquences symboliques

```
.  
|   doc  
|   |   build  
|   |   make.bat  
|   |   Makefile  
|   |   source  
|   notes.md  
|   README.md  
|   symbolic  
|       algorithmic.py  
|       discretize.py  
|       exceptions.py  
|       generator.py  
|       information.py  
|       __init__.py  
|       iosymb.py  
|       __pycache__  
|       recurrence.py  
|       sandbox  
|       sequence.py  
|       stochastic.py  
|       tests  
|       viz.py
```

Tests

Tests réalisés avec Doctest

```
algorithmic.py > lempel_ziv
118
119     Example :
120
121     >>> np.random.seed(9)
122     >>> a = np.random.choice([0,1],1000,replace=True, p=[0.7,0.3])
123     >>> A = S.Alphabet(['a','b'])
124     >>> seq = S.Sequence(a,A)
125     >>> lempel_ziv(seq)
126     0.6
127     """
128
129     try:
130         algorithm = eval(parsing)
PROBLÈMES   SORTIE   CONSOLE DE DÉBOGAGE   TERMINAL   JUPYTER
File "/home/lucas/scikits-symbolic/symbolic/algorithmic.py", line 125, in
    main__.lempel_ziv
Failed example:
    lempel_ziv(seq)
Expected:
    0.6
Got:
    0.6921676999572451
*****
1 items had failures:
    1 of   5 in __main__.lempel_ziv
***Test Failed*** 1 failures.
```

Figure – Exemple d'un test réalisé avec Doctest échoué volontairement

Documentation associée

Documentation avec Sphinx

```
class sequence.Alphabet(nsymb)
```

The set of states or symbols that can be visited for a sequence or Markov process realization.

Tests on State and Alphabet

```
>>> state1 = State('One')
>>> state1
State(- | One)
```

An integer representation of a state is only attributed once the state is inserted in an alphabet.

```
>>> state2 = State('Two')
>>> state3 = State('Three')
```

Alphabets can be created with a list of states:

```
>>> alpha = Alphabet([state1, state2, state3])
>>> alpha
Alphabet[State(0 | One), State(1 | Two), State(2 | Three)]
>>> print(alpha)
Alphabet[State(0 | One), State(1 | Two), State(2 | Three)]
>>> len(alpha)
3
```

Alphabets can also be created using only the length as argument.

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques

La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut

Spectre et filtre coupe-bande

Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique

Entropie

Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Interprétation des résultats

On s'intéresse à la valeur p ou "p-value" comme résultat du test-t. Celle-ci nous donne la probabilité de l'hypothèse nulle, i.e., que les deux populations sont statistiquement identiques. Lorsque l'on obtient une valeur p inférieure à 5%, $p \leq 0.05$, on peut considérer que la probabilité de l'hypothèse nulle correspond au hasard et donc que les deux échantillons sont statistiquement différents.

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques
La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut
Spectre et filtre coupe-bande
Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique
Entropie
Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word

Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Target word vs individual word

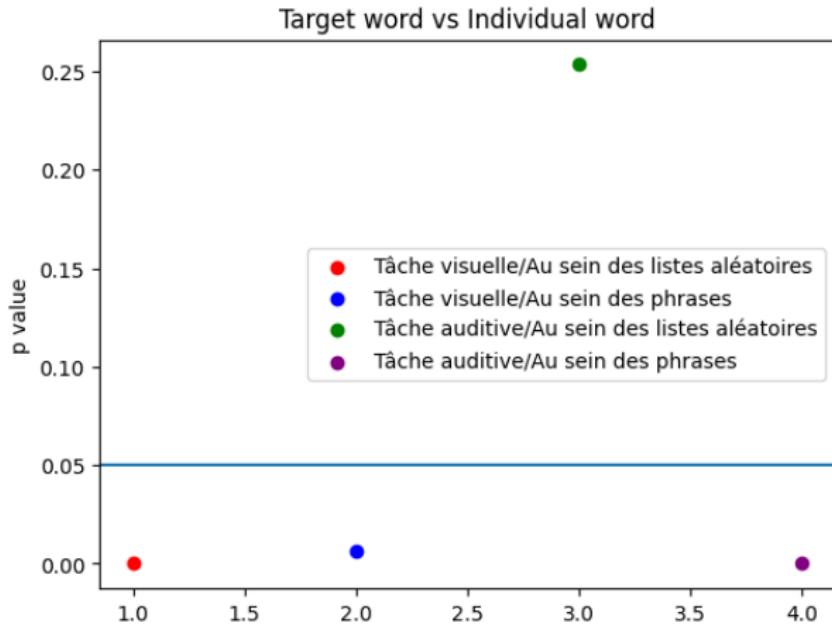


Figure – Tests t appariés entre les vecteurs d'entropie relatifs aux target words par rapport aux individual words

Notions de neurophysiologie

- De l'activité neuronale aux signaux neurophysiologiques
- La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

- Séries temporelles et filtre passe-haut
- Spectre et filtre coupe-bande
- Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

- Représentation symbolique d'un système dynamique
- Entropie
- Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

- Target word vs individual word
- Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Phrases simples vs listes aléatoires de mots issues de phrases simples



Figure – Tests t appariés entre les vecteurs d'entropie relatifs aux phrases par rapport aux listes aléatoires de mots

Notions de neurophysiologie

De l'activité neuronale aux signaux neurophysiologiques
La technique d'enregistrement MEG

Description du dataset

Pré-traitement des signaux MEG

Séries temporelles et filtre passe-haut
Spectre et filtre coupe-bande
Segmentation temporelle du signal

Analyse dans le cadre de théorie de l'information

Représentation symbolique d'un système dynamique
Entropie
Taux d'entropie et estimation

Développement et code Python

Résultats et tests statistiques

Target word vs individual word
Phrases simples vs listes aléatoires de mots issues de phrases simples

Tâche visuelle vs tâche auditive

Tâche visuelle vs tâche auditive

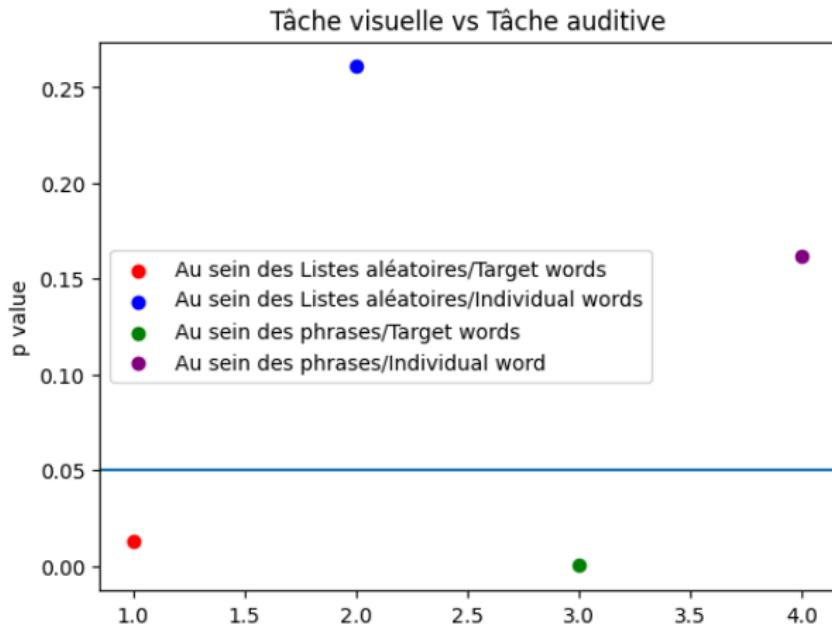


Figure – Tests t appariés entre les vecteurs d'entropie relatifs à la tâche visuelle par rapport par rapport à la tâche auditive

Conclusion et ouverture

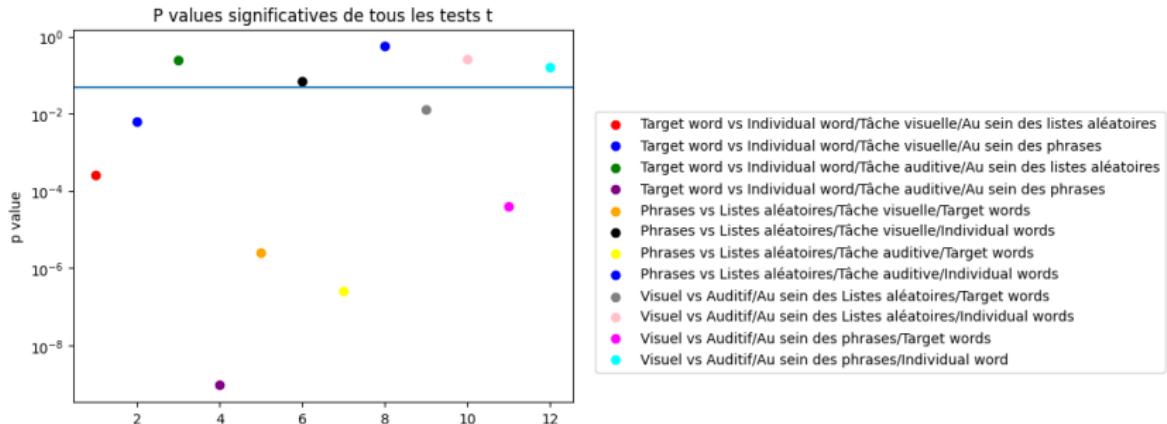


Figure – Valeurs p significatives des tests t. L'axe des ordonnées est en échelle logarithmique afin de pouvoir distinguer les différentes valeurs