

Cahier d'avancement

Data Visualisation

Université Lyon I

ASRI Amin
BOUBKER Ayoub
RUYTER Swann

Sommaire

<i>Sommaire</i>	2
<i>Rappels</i>	3
Thème - données personnelles	3
Problématique : Comment fournir une visualisation de données neurale issue de magnétoencéphalographie pour une première analyse post-processing?	3
Jeu de données : données neurales	3
<i>Thème(s) de la séance : planification du projet</i>	4
<i>Date : 17/11 à 30/11</i>	4
I. Recherches & cadrage	4
II. Réalisations attendues.....	5
III. Planification	5
<i>Thème(s) de la séance : recherches ciblées</i>	6
<i>Date : 30/11 à 9/12</i>	6
I. Réalisations attendues.....	7
II. Planification	7
Notes : difficultés dans le pré-traitement des données issues de matlab.	7
<i>Thème(s) de la séance : début du développement</i>	8
<i>Date : 9/12 à 14/12</i>	8
I. Réalisations attendues.....	8
II. Planification	9
<i>Annexe A</i>	13
<i>Peer-review projet Dataviz Lyon1 M2 2020-2021</i>	13
Date : 13/12/21	13
Groupe auteur de la relecture : 6	13
Groupe sur lequel porte la relecture : 10	13
Problème	13
État de l'art.....	13
Proposition	14

Rappels

Thème - données personnelles

Problématique : Comment fournir une visualisation de données neurale issue de magnétoencéphalographie pour une première analyse post-processing?

Jeu de données : données neurales

Document de cadrage du projet 10/12 (00h): [section cadrage](#)

Rendu final : <https://kraramel.github.io/TP-DATA-VIZ/>

Pour votre projet

- Utiliser un nom de repository github pertinent (pas ProjetDataviz13..)
- Images représentatives du projet, une grand format : teaser.png, une miniature (500x300) thumbnail.png ou gif à la racine de votre projet
- Mettre sur le site son nom + lien vers site web (ou emails)
- Lien vers le cours de visualisation, Université Lyon1, votre formation
- Sourcer et mettre des crédits: données, inspiration, etc.

S'assurer d'avoir un Readme.md à la racine de votre projet github, à minima:

- Version anglaise/résumé
- Un titre
- Une description
- + une image représentative du projet (.png ou .gif 500x300)

Pull request sur le site du cours, dossier 2021/gallery/

1. lire la documentation sur la façon de faire un [pull request](#)
2. Ajouter deux images représentatives du projet dans le dossier [gallery](#) en respectant le **format de nommage** des fichiers :
 - une grand format avec le nom : numerogroupe-teaser.png
 - une miniature (500x300) : numerogroupe-thumbnail.png ou numerogroupe-gif
3. Ajouter un fichier de description simple décrivant le projet en français numerogroupe-desc-fr.md et un autre en anglais numerogroupe-desc-en.md, en respectant les modèles suivants (titre, lien et description) [numeroprojet-desc-fr.md](#), [numeroprojet-desc-en.md](#)

Le projet constituera 60 % de la note de l'UE, avec la répartition suivante:

- critique visu (10%)
- cadrage et document de suivi (10%)
- peer review (5%)
- présentation et démo (35%)
- réalisation technique (40%)

Thème(s) de la séance : planification du projet

Date : 17/11 → 30/11

I. Recherches & cadrage

Quel est le problème abordé / à quel besoin répondez vous ?

Le problème abordé est la visualisation des données neurales. En effet, les jeux de données issus d'imagerie cérébrale demandent un certain niveau d'expertise (canaux ou électrodes, fréquence, modalités, lien avec les données comportementales).

Ce projet répond à la difficulté d'obtenir des visualisations claires post-processing, c'est-à-dire lorsque les données ont été nettoyées et que l'on souhaite obtenir une première visualisation de l'activité cérébrale.

Ici, l'objectif principal est de fournir dans un premier temps une analyse de la connectivité des différents canaux.

À qui s'adresse la visualisation, quelles tâches seront effectuées au travers de votre projet ?

Clarifier le public principal de votre projet. Lister 3 tâches que votre projet permettra d'effectuer. Par ex. : "identifier l'application qui consomme le plus de données sur mon smartphone". Expliquer pourquoi ces tâches vous semblent les plus importantes.

Le public ciblé est un public très spécifique, potentiellement des étudiants ou jeunes chercheurs voulant faire une première analyse générale sur des données neurales issues de magnéto-encéphalographie (MEG).

Tâche 1 : fournir une heatmap de l'activité cérébrale selon les différents essais et selon une bande de fréquence

Tâche 2 : fournir une visualisation de la connectivité fonctionnelle vis-à-vis des différents canaux MEG (en graphe sur une map crânienne, all-to-all connectivity)

Tâche 3 : fournir une visualisation de la connectivité fonctionnelle en graphe circulaire.

Sources de données choisies

Lister les sources de données identifiées, pour chacune identifier leur intérêt principal et leurs limites potentielles. Clarifier si vous souhaitez en fusionner plusieurs ou pensez privilégier une en particulier. En cas de surprise/problème imprévu lié aux données, quel est votre plan de secours ?

Les données sont composées pour chaque participant, de 144 essais eux-mêmes composés de 246 canaux MEG sur 100 fréquences. L'intérêt principal est d'identifier des patterns d'activation cérébrale (principalement dans le cortex auditif et moteur) en fonction de la modalité (i.e., 3 modalités en 144 essais, soit 48 essais par modalité).

La principale limite réside dans la grande quantité de données : $144 * 246 * 100 * 30$. Ainsi, nous pensons ne pas utiliser entièrement le jeu de données afin de simplifier l'analyse. En ce sens, nous nous concentrerons sur une bande de fréquence, et peut être pour un seul participant. Ces contraintes sont à définir selon les difficultés que nous allons rencontrer.

Un autre enjeux réside dans la façon dont nous allons croiser les différents jeux de données. En effet, le package MNE de python est conçu pour cela, nous ne savons pas si nous arriverons exactement à reproduire ces résultats avec D3.

En cas d'imprévu, nous changerons de jeu de données et utiliserons des données issues de nos activités sur les réseaux sociaux.

Travaux important liés au projet

1. Contexte

Types de questions : dans une approche exploratoire, comment fournir une visualisation simple de l'activité cérébrale ? De la connectivité des canaux/électrodes ?

Comment fournir une visualisation permettant une jonction entre les données neurales et comportementales ?

Comment repérer rapidement des outliers ?

2. Brainstorming : garder les meilleures questions

3. Esquisser les données (type de format...)

4. Cibler : Crazy 8s

II. Réalisations attendues

✓ Constituer le groupe

✓ Thématique définie : analyse de connectivités neurale

✓ Repository Github

✓ Premières esquisses

III. Planification

<i>Tâche</i>	<i>Priorité</i>	<i>Description</i>	<i>Temps alloué</i>	<i>Assignment</i>
1	1	Récupérer les données	20mn	Swann
2	1	Création repository github	20mn	Amin
3	2	Rédaction du cadrage	1h30	Swann
4	2	Rédaction du cadrage	1h30	Amin
5	2	Rédaction du cadrage	1h30	Ayoub
6	3	Etat de l'art	1h30	
7	3	Etat de l'art	1h30	
8	3	Etat de l'art	1h30	
9				

Thème(s) de la séance : recherches ciblées

Date : 30/11 → 9/12

I. Recherches & cadrage

Travaux important liés au projet

Lister 3 projets liés au votre, par exemple : des projets avec des jeux de données similaires, des projets proposant des techniques de visualisation que vous trouvez intéressantes, des articles scientifiques présentant une technique ou une description du problème qui vous sera utile... Pour chaque travail, expliquer en quoi il vous semble intéressant par rapport à votre projet, et en quoi il pourrait être amélioré.

A Tool for Interactive Data Visualization: Application to Over 10,000 Brain Imaging and Phantom MRI Data Sets

https://www.researchgate.net/publication/298429579_A_Tool_for_Interactive_Data_Visualization_Application_to_Over_10000_Brain_Imaging_and_Phantom_MRI_Data_Sets

Seul article trouvé croisant D3 et l'imagerie cérébrale. Cela peut par conséquent donner de bonnes pistes à approfondir.

Self-portraits of the brain: cognitive science, data visualization, and communicating brain structure and function

https://cns.iu.edu/docs/publications/2015-goldstone-self-portraits.pdf?fbclid=IwAR0jn9Bq_9eWC3mzgLF4Pp2kcoHG_j_9V_bS7rBtHjpScmULL3GAX6gbXnk

Ensemble de visualisations possibles, peut donner de bonne idées.

https://brainhack.sparkboard.com/project/5c8fb2d88e310700242ba3a5?fbclid=IwAR0CVTWQfiXcRikrHRLy-LJv_-OFDkBWtd_ZUHa8XFGVfpNUr-rFvZi5YTI

Seul travail montrant un résultat de visualisation d'électrodes cérébrales avec D3.

Organisation

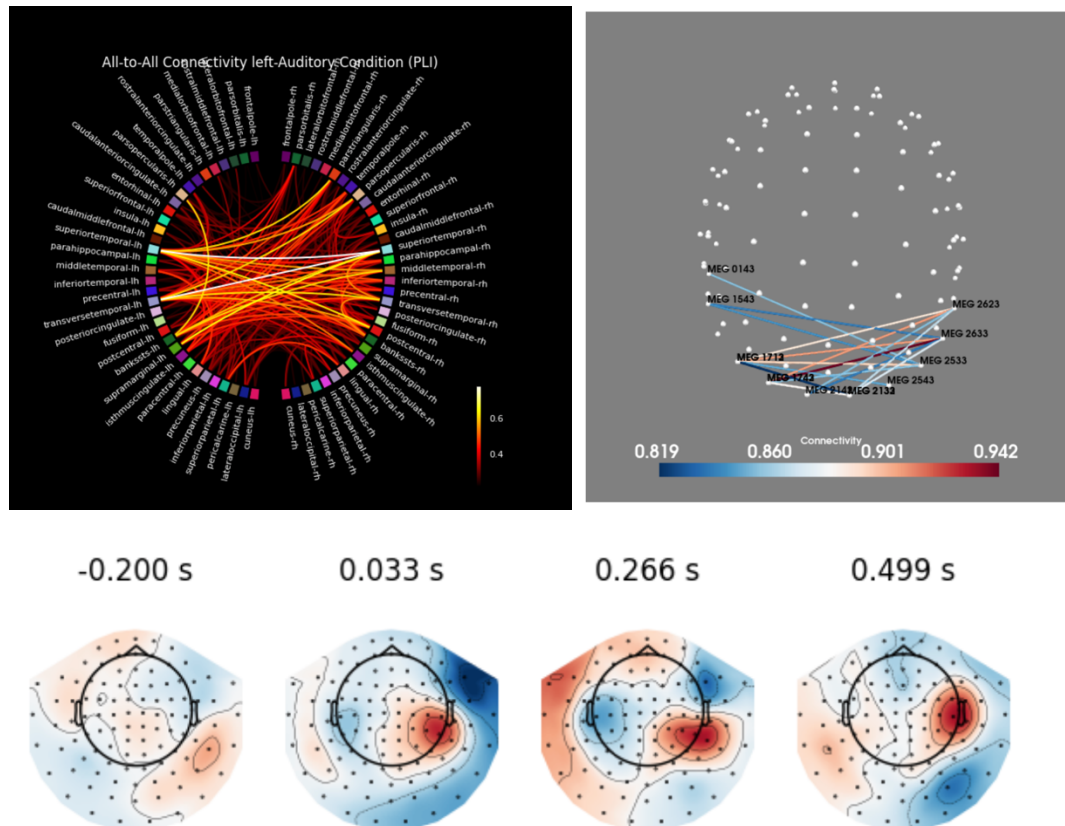
Nous avons mis en place un groupe facebook où nous pouvons rapidement converser. Lié à cela, nous avons un répertoire github afin de centraliser les réalisations. Enfin, nous avons ce document de suivi partagé sous Word.

Nous avons prévu de réaliser 75% du projet durant les vacances de Noël. Pour ce faire, nous avons réparti les tâches, que nous allons précisément définir le premier weekend des vacances de Noël. En ce sens, chacun aura des objectifs spécifiques à atteindre, et nous avons réservé au minimum 3 demi journées durant les vacances de Noël pour mettre en commun notre travail.

Pour le moment, le pré-traitement des données sera réalisé par Swann, le design est principalement attribué à Ayoub, et Amin s'occupe de la conception du développement (des différentes étapes, modules, etc.) qui seront réalisés par chacun de nous.

Scan des esquisses finales

Etant donné la complexité de représenter un cerveau, nous préférons fournir des captures d'écran.



I. Réalisations attendues

- ✓ Peer-to-peer review : cf. [annexe A](#)
- ✓ Plan d'organisation pour les semaines à venir

II. Planification

Tâche	Priorité	Description	Temps alloué	Assignment
1	1	Pré-traitement des données		Swann
2	1	Recherche de projets similaires		Amin & Ayoub

Notes : difficultés dans le pré-traitement des données issues de matlab.

Thème(s) de la séance : début du développement

Date : 9/12 → 14/12

I. Recherches & cadrage

L'objectif est d'explorer ces différentes visualisations afin d'en extraire ce qui est pertinent. Pour ce faire, lorsque l'on explore ces projets, l'on prendra des notes afin de permettre à l'ensemble des membres du groupe d'avoir les informations nécessaires et de ne pas avoir à refaire le travail.

Recherche de projets similaires :

https://github.com/wchunchou/brain-map?fbclid=IwAR1iI1YPEXl-oF9Eaziu_wP8LkTVxEhe_-T_d17xJPSbog1CDC9ynTR-s7Q

(notes)

<https://github.com/bricaud/brainviz2?fbclid=IwAR1UnRPjlkYmfq4SBMQSqHDjKpvp7PGUy6dV6NhLM-HejVLx-7CuPnqZxRU>

(notes)

https://github.com/rpassas/Visual_Brain?fbclid=IwAR2JeUfkpyDxxvpEEpxYkC9F2_8gJUuXK3npL7rbjwmjg15t0A6CINCXLjk

(notes)

https://github.com/greallye/brain-stem-eeg?fbclid=IwAR11tji31kkevcsM_6hiMH52-tP0fO0Jjj-n6AA8B2SDJKp_iDxyy5dSY1c

(notes)

<https://github.com/mjain72/brainwave?fbclid=IwAR1mK1b2So8pSkhLnHr2rYPx5V7FQxM9dYMjiMBw-vRnBt3JQw9RWjLr2cE>

(notes)

https://github.com/mne-tools/mne-workshops/tree/master/2018_06_Amsterdam?fbclid=IwAR0S19CEalyV_wIvmFW3bH_aWBdYhj8xa5K7MJ0R9tsBKCJD9LXdIQWd_Y8

(notes)

<https://www.frontiersin.org/articles/10.3389/fninf.2016.00009/full?fbclid=IwAR0BsbHUB0XDYAGocWX6KINx9LWdLyOYDD6vPGl-UtGLmlc7uO11YYAWjZ0>

(notes)

https://github.com/BrainJS/brain.js?fbclid=IwAR2JeUfkpyDxxvpEEpxYkC9F2_8gJUuXK3npL7rbjwmjg15t0A6CINCXLjk

(notes)

I. Réalisations attendues

✓ Peer-to-peer review : cf. [annexe A](#)

II. Planification

<i>Tâche</i>	<i>Priorité</i>	<i>Description</i>	<i>Temps alloué</i>	<i>Assignment</i>
1	1	Pre-traitement des données		Swann
2	1	Analyse des sources		Amin & Ayoub
3	2	Recherche de données théoriques en attendant données réelles		

Thème(s) de la séance : acquisition de l'ensemble des données et généralisation

Date : 14/12 → 28/12

I. Réalisations attendues

✓ Obtenir les données exploitables

✗ Avoir une base significative de visualisations possibles et leurs codes associés

II. Planification

Tâche	Priorité	Description	Temps alloué	Assignment
1	1	Extractions de données	2h	Swann
2	2	Développement		Ayoub & Amin

Notes : difficultés dans l'extraction de la localisation des canaux.

Thème(s) de la séance : Soutenance des projets

Date : 28/12 → 12/01

I. Réalisations attendues

- ✓ Soutenance
- ✓ Données exploitables
- ✓ Visualisations réalisées
- ✓ Description du projet

II. Planification

Tâche	Priorité	Description	Temps alloué	Assignment
1	1	Développement		Ayoub & Amin
2	1	Rédaction du projet		Swann
3	1	Préparation de l'oral		Swann
4				
5				
6				
7				
8				
9				

Notes :

Thème(s) de la séance : Validation des changements

Date : 12/01 → 19/01 // Rendu final : 21/01 00h

Réalisations attendues

✓ Must-do

✓ Nice-do

III. Planification

Tâche	Priorité	Description	Temps alloué	Assignment
1	1	Reprise du code		Ayoub & Amin
2	1	Rédaction readme		Swann

Notes :

Annexe A

Peer-review projet Dataviz Lyon1 M2 2020-2021

Date : 13/12/21

Groupe auteur de la relecture : 6

Groupe sur lequel porte la relecture : 10

Problème

Reformuler le problème tel que vous l'avez compris (2/3 phrases maximum).

Le problème consiste à analyser les habitudes de jeu au travers de différents indicateurs : le temps de jeu selon la période qui peut être en période d'examen ou non, et durant la semaine donc entre une période cours / « repos » ; ainsi qu'une classification des jeux selon des catégories.

Clarifier si le problème vous semble valide. Si oui proposer une direction en particulier qui vous semble intéressante à creuser, si non proposer une alternative réaliste (même données / même contexte global).

L'objectif principal, de ce que nous en avons compris, est de mieux gérer son temps de jeu par analyse de l'activité sur une période / en fonction des jeux. L'approche nous semble pertinente, néanmoins, comment pouvoir faire une analyse sans éléments de comparaisons ? Par exemple, je me fixe un objectif de 5h maximal de jeu par semaine : combien de temps y ai-je passé effectivement ? Si j'ai dépassé ce temps maximal, quelles sont les périodes sur lesquelles je pourrais diminuer ce temps ? Y a-t-il des jours / moments où je ne peux pas me permettre de jouer à la suite de contraintes extérieures (p.ex., charge de travail importante) ?

En effet, suite aux éléments que nous avons à notre connaissance, cette visualisation semble ne fournir qu'une analyse descriptive, sans pouvoir réellement en tirer des conclusions. Par exemple :

Est-ce que les tendances générales de jeu vidéo impactent le temps de jeu du joueur ?

Si oui, sur quelle période de temps (p.ex., va augmenter son temps de jeu pendant 5 jours)

Si oui, les tendances générales de jeu sont-elles cohérentes avec les préférences des joueurs ?

De plus, il serait pertinent de fournir une classification par « type » de jeu : en ligne, hors ligne, jeu de coopération ou de compétition, et voir le comportement du joueur en fonction de cela. Cela pourrait permettre, par exemple, de savoir que je passe beaucoup de temps sur les jeux en compétition et peu en coopération, donc en période d'examens je vais m'interdire les jeux en compétition.

État de l'art

Proposer deux références (projet, site, article scientifique) complétant celles citées, soyez précis et expliquer pourquoi elle vous semble intéressante pour le groupe

1. <https://www.githyp.com/grand-theft-auto-v-101852/?tab=player-count>

Donne un élément comparatif : l'activité des joueurs en général sur ce jeu à comparer avec l'activité du joueur en question.

2. <https://steamdb.info>

Est une alternative à la ressource citée, fournissant potentiellement des éléments complémentaires.

Proposition

Expliquer ce que vous aimez (3 phrases).

Pour un amateur de jeu vidéo, il est intéressant d'obtenir une visualisation claire de son activité : par exemple, cela peut permettre de se rendre réellement compte de son activité, comme le temps d'écran, permettant de quantifier la différence entre « avoir l'impression » et « avoir réellement fait ». Ainsi, si j'avais décidé de ne plus jouer plus de 1h le soir, et que je vois que j'y ai passé 3h, cela peut m'aider à rectifier mon comportement.

Expliquer ce que vous auriez fait différemment (3 suggestions).

Nous aurions suggéré ces visualisations à des professionnels des jeux vidéo, ou des personnes addictes aux jeux (population cible plus restreintes avec des objectifs clairement définis).

Dans le premier cas, pour que les professionnels puissent améliorer leur efficacité/efficience dans les moments de jeu comparés à leurs performances, et dans le second cas pour accompagner les personnes addictes aux jeux à réduire leur activité (ce sont deux alternatives possibles). L'ajout de la composante âge du joueur apparaît aussi pertinente dans la mesure où l'activité serait différente. Fournir la possibilité de visualiser des données sur une période plus importante.

D'une façon générale, nous aurions cibler davantage le public, et cibler davantage l'intérêt de la visualisation (à quel besoin répond-elle réellement ?) et enfin de pouvoir inscrire cette visualisation dans un contexte plus large pour dégager des tendances à long terme (par exemple, croiser l'humeur/temps de jeu, performances/moment de jeu).

Expliquer ce que vous aimeriez voir (3 suggestions).

Vis-à-vis du temps, ce serait intéressant de pouvoir voir le temps de jeu par rapport au temps de cuisine, d'université, de révisions, de relations sociales ou encore du sport, sommeil pour pouvoir quantifier l'importance de ce temps dans la vie du joueur.

Par ailleurs, nous aimerions voir une visualisation du temps passé sur le jeu et la performance du joueur. Comme ça, nous pourrions voir si le temps de jeu influe sur la performance et oui, dans quelle mesure.

Enfin, ce serait intéressant de voir davantage de rapport au subjectif : un temps de jeu croisé à l'humeur ou encore le niveau de fatigue, pour répondre à la question de ce qu'apporte le jeu ? Des émotions positives ? Négatives ? Est-ce que ce temps empiète sur le sommeil ?