

## RAPPORT FINAL

PSC HSS 02

---

# Intégration collective des phénomènes d'attention

---

*Encadré par :*

Jérôme Sackur  
Maître de conférences ENS,  
LSCP ENS-CNRS-EHESS

*Proposé par :*

Aznag Abdellah  
Moal Candice  
Ravetta Antoine  
Sainvet Hector  
Truffinet Olivier

*Remerciements*

En premier lieu, nous tenons à remercier M. Jérôme Sackur, le tuteur de notre projet, pour ses conseils avisés tout au long de notre travail et son optimisme indéfectible lors de nos entretiens.

Nous remercions également Mme Anne Dulphy et M. Jean-Yves Grenier, les coordinateurs du département d'Humanités et Sciences Sociales, pour avoir permis le lancement de notre projet sur une question aussi ambitieuse que l'attention des élèves.

L'ensemble des expériences n'aurait pas été possible sans le concours des professeurs volontaires pour nous accueillir : Mme Laurence Grimaud et M. Vincent Martigny, et bien sûr les élèves de leurs classes de chimie et de sciences politiques. Nous remercions aussi M. Manuel Joffre, professeur de physique, pour nous avoir permis d'accéder à des boîtiers de mesure plus commodes que les téléphones, mais malheureusement peu exploités.

Enfin, nous tenons à remercier les nombreux interlocuteurs et interlocutrices qui ont contribué à ce projet comme Mme Latifa Berkous, de la direction ingénieur pédagogique de l'Ecole polytechnique, pour avoir soutenu notre ambition ; MM. Thierry Bodineau et Damien Rohmer, professeurs de mathématiques appliquées et d'informatique, pour leurs conseils sur les méthodes d'analyse ; M. Tadas Baltruitas, pour avoir développé le logiciel OpenFace de traitement vidéo ; et les nombreux professeurs et chefs de cabinet de rectorats qui ont répondu à notre sondage ou contribué à sa diffusion.

*Image de couverture : Séance de chimie du 21/12 traitée par OpenFace en Image Multi-visage*

## Résumé

Une grande partie de l'apprentissage des élèves se joue dans la classe. Lorsque le ou la professeur donne cours, les élèves peuvent toutefois être attentifs à différents degrés. Or, leur niveau d'attention engage la qualité et l'efficacité de cet apprentissage. Dans cette étude, nous cherchons à développer une démarche de mesure de l'attention des élèves en classe, à la fois peu intrusive et facilement réalisable avec les moyens d'un projet mené par des étudiants.

Nous avons exploré des méthodes d'analyse manuelle qui consistent à observer et faire noter les élèves par un évaluateur extérieur. Ces méthodes sont fiables et souples mais néanmoins assez lourdes et peu reproductibles. Nous développons alors un outil d'analyse automatique et informatisé d'une vidéo de la classe. La *ground truth* (véritable état psychique, inaccessible à l'observation) est établie par un système d'auto-évaluation de la part des élèves.

Pour concevoir ces expériences et envisager l'ensemble des facteurs potentiellement pertinents à prendre en compte, nous sondons les professeurs des académies d'Amiens et de Bordeaux. L'enjeu des réponses dépasse largement les moyens atteignables par notre démarche de mesure mais les réponses mettent en évidence la pertinence de certains facteurs (fatigue, faim, nombre d'élèves) ou non (sourire, nombre d'enseignants).

D'une part, nous avons essayé d'exhiber quelques profils d'évolution de l'attention (évolution temporelle et comportementale, corrélation des gestes avec le niveau d'attention global). D'autre part, nous avons largement entamé le développement d'un outil d'analyse automatique (choix du logiciel d'analyse vidéo, extraction des données brutes, nettoyage et exploitation en variables physiques) dont les données nourrissent un embryon d'analyse suivant deux méthodes : analyse de variance et quasi-implications.

## Table des matières

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction et objectifs</b>  | <b>4</b>  |
| 1.1      | Introduction . . . . .  | 4         |
| 1.2      | Objectifs . . . . .   | 5         |
| <b>2</b> | <b>Degré de réalisation des objectifs</b>                                     | <b>6</b>  |
| 2.1      | La notion d'attention . . . . .   | 6         |
| 2.2      | Protocole et qualité des mesures . . . . .                                    | 6         |
| 2.2.1    | Protocole et variables mesurées . . . . .                                     | 6         |
| 2.2.2    | Protocole de traitement . . . . .   | 8         |
| 2.3      | Développement de la méthode d'analyse . . . . .                               | 9         |
| 2.3.1    | Analyse manuelle . . . . .  | 9         |
| 2.3.2    | Analyse informatisée et automatique . . . . .                                 | 9         |
| 2.3.3    | Intégrer une dimension d'étude de terrains auprès des professeurs .           | 14        |
| <b>3</b> | <b>Résultats finaux</b>   | <b>15</b> |
| 3.1      | Résultats expérimentaux et corrélation . . . . .                              | 15        |
| 3.1.1    | Expérience prototype évaluée . . . . .  | 15        |
| 3.1.2    | Expérience longue . . . . .   | 16        |
| 3.2      | Modèle et analyse . . . . .   | 19        |
| 3.2.1    | Analyse des réponses au sondage seules . . . . .                              | 19        |
| 3.2.2    | Analyse couplée des réponses et des vidéos . . . . .                          | 25        |
| <b>4</b> | <b>Croisement au témoignage empirique des professeurs</b>                     | <b>32</b> |
| 4.1      | L'intuition d'un savoir-faire ou plutôt d'un <i>savoir-détecter</i> . . . . . | 32        |
| 4.2      | Elaboration du sondage . . . . .  | 32        |
| 4.3      | Résultats . . . . .   | 33        |
| 4.4      | Conclusion . . . . .  | 36        |
| <b>5</b> | <b>Discussion du projet : apports et limites</b>                              | <b>37</b> |
| 5.1      | Apports du projet . . . . .   | 37        |
| 5.2      | Limites des résultats, et prolongements possibles . . . . .                   | 37        |
| 5.3      | Discussion sur les outils et protocoles utilisés . . . . .                    | 38        |
| <b>6</b> | <b>La vie de projet collectif</b>   | <b>40</b> |
| 6.1      | L'organisation en groupe . . . . .  | 40        |
| 6.2      | La coopération et la complémentarité . . . . .                                | 40        |
| <b>A</b> | <b>Modélisation mathématique en chaînes de Markov</b>                         |           |

# 1 Introduction et objectifs

## 1.1 Introduction

En septembre 2018, un élève chinois découvre sur Twitter une photo d'une classe de son école avec en surimpression "ID : 000010, State 1 : Focused," une ligne de texte au-dessus d'un rectangle vert entourant le visage d'un de ses camarades [7].

En fait, ce sont 28 000 élèves de 6 écoles différentes qui ont testé, sans le savoir, un système d'évaluation automatique. Ce projet veut mettre la reconnaissance faciale au service de l'éducation en augmentant la concentration des élèves, et en espérant ainsi améliorer leur capacité à apprendre. L'expérience "Class Care" s'inscrit en fait dans le grand plan de développement de l'intelligence artificielle du gouvernement chinois (Next Generation Artificial Intelligence Development Plan (NGAIDP)). La hauteur des moyens engagés par le gouvernement chinois pour développer ce type d'outil donne la mesure de l'enjeu que représente l'étude de l'attention en classe.

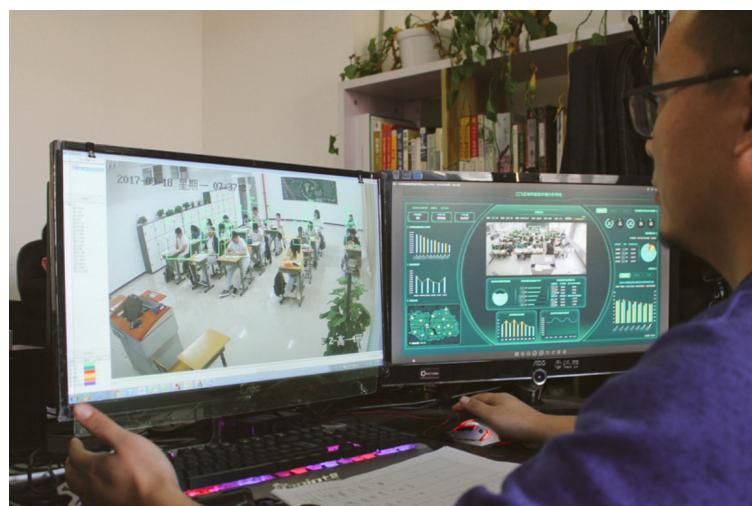


FIGURE 1 – Un programmeur du centre de recherche en éducation Hanwang présente le système "Class Care" à Beijing le 2 janvier 2019. Xue Yujie/Sixth Tone

Lorsque notre projet débuta, nous n'avions pas connaissance des ambitieuses expériences réalisées en Chine – et qui posent certaines questions éthiques. Néanmoins, nous imaginions bien l'intérêt d'un outil de détection automatique en temps réel de l'attention des élèves. A moyens termes, il serait un outil de mesure efficace pour des recherches scientifiques sur l'attention en classe. C'est dans cette perspective que nous avons inscrit notre démarche : mener en intégralité un travail de détection de l'attention.

## 1.2 Objectifs

Suite à notre rapport intermédiaire, nous voulons rappeler ici l'évolution de nos objectifs. L'idée initiale était d'utiliser des outils informatiques et vidéo existants pour étudier et modéliser des phénomènes collectifs d'attention (ex : perturbations dues aux voisins, zones de déconcentration), puis éventuellement utiliser cette modélisation pour améliorer la concentration d'une classe (ex : en modifiant le plan de classe).

Cependant, nous avons vite constaté qu'il n'existe actuellement pas d'outil performant et accessible capable de détecter sur une simple vidéo si des individus sont attentifs ou non. Quelques travaux ([3]) tentent d'effectuer une telle détection à partir de critères visuels (ex : direction du regard, mouvements faciaux), mais n'arrivent pas à corrélérer clairement ces critères avec le ressenti des personnes sur leur propre état d'attention, la *ground truth* de toute étude de ce type.

Nous avons donc décidé de nous concentrer sur ce sujet, et d'étudier si des méthodes d'analyse de vidéos **appliquées à une classe entière** permettaient de prédire l'état d'attention d'élèves, dans l'idée de développer un outil **réaliste** capable d'effectuer cette tâche à l'aide d'une simple caméra installée dans la salle de cours. La distinction individuel/collectif a ici son importance, car la plupart des logiciels de détection de visages nécessaires à ce traitement peinent à reconnaître plusieurs visages simultanément de manière stable dans le temps.

Parallèlement, dans la lignée de notre objectif initial, nous avons effectué quelques analyses statistiques simples sur nos données **manuelles** (réponses des élèves sur leur propre état d'attention et analyse des vidéos "à la main") pour répondre à certaines questions basiques (L'état d'attention d'un élève est-il fortement influencé par celui de ses voisins ? Par l'état général de la classe ?). Les résultats de cette étude seront présentés en partie 3.1.

## 2 Degré de réalisation des objectifs

### 2.1 La notion d'attention

Notre travail bibliographique s'est achevé véritablement au courant du mois de janvier avec la découverte des travaux de Mirko Raka. [3] [4] Nous ne prétendons pas avoir épuisé l'état de l'art dans notre domaine, l'étude de l'attention des élèves dans une classe. Néanmoins, les difficultés à cerner un domaine de recherche du même type d'échelle que celui que nous ambitionnions (filmer une classe depuis l'extérieur avec un questionnaire subjectif) attestent de la multiplicité des approches des études sur l'attention et *a fortiori* de l'attention elle-même. En effet, le concept *attention* est employé aussi bien dans des études neurophysiologiques qui mesurent l'activité électrique du cerveau que dans des études comportementales qui mesurent le temps de fixation du regard.

Or, le cœur de notre démarche scientifique consiste d'abord à trouver des critères pertinents pour qualifier l'*attention* d'un élève, ou simplement reconnaître qu'un élève est *attentif*. Nous en sommes arrivés à une approche mixte qui se rapproche de la notion d'*alertness* utilisée en sciences cognitives, ou encore de concentration. L'attention selon nos critères se trouvera à l'intersection entre des **données objectives** (variables physiques du corps), un **ressenti intérieur** (appréciation humaine subjective intérieure) et un **ressenti extérieur** (appréciation humaine subjective extérieure). La notion est donc floue conceptuellement mais construite pour être scientifiquement exploitable.

### 2.2 Protocole et qualité des mesures

#### 2.2.1 Protocole et variables mesurées

##### *Variables mesurées*

Le protocole cherche à collecter les 3 types de variables précédemment établies :

- **Variables objectives** : vidéo filmée en plan de classe fixe et analysée automatiquement par ordinateur. On en retire des informations sur le regard de l'élève, ses expressions faciales, ses mouvements, etc.
- **Variables subjectives internes** : sondages individuels sur téléphone. On peut demander aux élèves d'évaluer leur attention de manière quantitative (note sur 10) ou qualitative (choix entre un ensemble d'états mentaux).  
Nous récoltons aussi des informations sur l'état général de l'élève lors d'une séance donnée : fatigue, faim, intérêt pour le cours, affinités avec ses voisins.
- **Variables subjectives externes** : un observateur extérieur peut donner un score de 1 à 10 à l'attention d'un élève, ou relever son action présente (prendre des notes, parler à son voisin, etc).

*Méthode et objectifs : film et réponses par téléphone*

Afin de mener une étude à long terme, nous avons réalisé l'expérience longue en partenariat avec la classe de CHI 421, après avoir testé nos protocoles dans les cours de CHI 431 et HSS 414b. En CHI 431, nous avions filmé les élèves pendant deux séances et leur avions demandé de noter leur attention sur une échelle linéaire. Or, l'échelle linéaire peut-être interprétée subjectivement de manière très différente par les élèves et ainsi engendrer des données confuses. En HSS 414b, nous avons soumis aux élèves des questionnaires d'auto-évaluation (logiciel Wooclap) grâce auxquels ils pouvaient renseigner leur état parmi cinq possibles : attentif, distracté par l'environnement, rêveur, esprit vide, distracté par l'expérience (pentagone des états). En CHI 421, nous filmions une séance entière (90 min.) tout en proposant aux élèves de renseigner leur état mental sur leur téléphone (Wooclap) toutes les 12 minutes.

| Expérience        | CHI 431         | HSS 414b                | CHI 421             |
|-------------------|-----------------|-------------------------|---------------------|
| Objectifs         | Protocole vidéo | Protocole questionnaire | Expérience longue   |
| Nb d'élèves       | 30              | 18                      | 17                  |
| Format sondage    | Linéaire        | Pentagone des états     | Pentagone des états |
| Fréquence sondage | 10 min.         | 30 min.                 | 12 min.             |

*Structures spatiales de la classe*

Pour étudier les **structures spatiales** de l'attention, on s'attache aussi à relever la position des élèves dans la classe. Pour cela, les places étaient numérotées et les élèves devaient s'identifier par leur numéro en début de séance. Malheureusement, la correspondance entre un élève, sa place réelle et le numéro enregistré est parfois trop fluctuante pour être exploitée : cela nous a obligé à identifier les élèves à partir de leurs temps de réponse observés sur la vidéo (procédé assez fastidieux). Nous aurions pu établir un plan de classe fixe ou numérotter les places plus clairement.

*Variations temporelles*

Pour étudier les **variations temporelles** des états d'attention, il est nécessaire de mesurer l'heure de la réponse du sondage. Cependant, le logiciel Wooclap ne permet pas d'avoir accès à l'instant de la réponse de l'élève. Nous demandions donc aux élèves de renseigner l'heure de leur réponse. Parallèlement, nous avons aussi développé un second outil qui utilise les boîtiers TurningPoint que possède l'école. Après avoir discuté avec M. Manuel Joffre, qui utilise ces boîtiers pour son cours de physique, et avec l'entreprise Turning Point **nous avons codé et testé un programme qui permet d'enregistrer les réponses des élèves ainsi que les temps de réponse précis associés**. Malheureusement, faire appel à la société Turning Point pour acquérir les licences d'utilisation des boîtiers a occasionné de longues démarches de négociation et de recherche de caution. Finalement, nous sommes rentrés en possession de cette technologie trop tard (Février 2019) pour l'utiliser dans l'ensemble de notre expérience. Mais nous espérons que ces boîtiers et le

code serviront à tout autre projet étudiant similaire.

### 2.2.2 Protocole de traitement

Pour mesurer les **variables objectives**, les vidéos sont analysées par le logiciel libre de traitement de vidéos OpenFace [6]. Ce logiciel traite un fichier .mp4 et retourne un fichier .csv, qui associe à un visage reconnu un identifiant et un ensemble de variables de position, d'expressions faciales, etc. Nous avons exploité les variables suivantes : AUs<sup>1</sup>, ligne de vue de chaque oeil (uniquement pour étudier la focalisation des deux yeux), direction du regard (moyennée sur les deux yeux et exprimée en termes de longitude et latitude plutôt qu'en vecteur 3D), angles de rotation de la tête, position de la tête.

Voici le protocole pour obtenir les données :

1. Fusionner les extraits bruts de vidéos (*rushs*) qui durent 15 min. en une longue vidéo de 90min.
2. Découper spatialement (*cropper*) la longue vidéo en zones contenant au plus **4 visages** car OpenFace ne peut pas détecter plus de 4 visages en même temps sur une vidéo.
3. Découper temporellement chaque zone extraite (*crop*) autour des réponses Wooclap.<sup>2</sup> On conserve des vidéos d'environ 30 secondes *centrées à une minute avant la réponse Wooclap*, pour éviter l'écueil évident d'analyser les attributs physiques de l'élève au moment où il est penché sur sa table et écrit sur son portable.
4. Traiter chaque zone extraite sous OpenFace. On obtient environ 6 fichiers .csv bruts par zone, donc 30 fichiers .csv pour une seule séance.
5. Extraire de chaque fichier excel brut les données nécessaires grâce à un programme informatique VBA.

Parallèlement, nous traitons les réponses au sondage Wooclap comme les **variables internes** par des traitements statistiques classiques.

Pour les **variables externes**, un observateur note les vidéos en regardant une à une les zones autour des réponses.

---

1. AUs désigne les Action Units, un système de critères universels de détection des mouvements du visage via entre autres la contraction des muscles

2. En traitant 90min. de vidéo d'un coup avec OpenFace, nous obtenons un fichier excel de 1,5Go (!) beaucoup trop volumineux pour être exploitable.

## 2.3 Développement de la méthode d'analyse

### 2.3.1 Analyse manuelle

Deux analyses manuelles ont été menées : chiffrer l'attention et relever les actions. Ces deux méthodes répondent à deux démarches très distinctes.

**Chiffrer** consiste à donner une note de 1 à 10 de l'état d'attention d'un élève. Cette note serait un substitut ou un complément à la note du sondage individuel fait par l'élève lui-même. La méthode appartient à une *démarche subjective*. (cf. 3.1.1)

**Relever manuellement** les actions d'un élève consiste à substituer ou compléter les données physiques objectives obtenues par traitement informatique.

Un observateur humain extérieur semblerait a priori capable d'évaluer l'attention d'un élève par des critères subjectifs mais fiables. Un professeur dans sa classe effectue sans doute le même type d'évaluations conscientes ou inconscientes pour mesurer l'attention de sa classe (cf. les résultats du sondage 4.3). Mais, évaluer de l'extérieur n'apporte jamais une certitude quant à ce qui se passe *réellement* dans la tête de l'élève. Par exemple, il est apparu difficile de comprendre un élève qui ne bouge pas et fixe constamment devant lui (absorbé ou absent ?). La tâche est donc longue et probablement biaisée. (cf. 3.2.2)

### 2.3.2 Analyse informatisée et automatique

Le traitement informatique repose d'abord sur le traitement par le logiciel libre OpenFace. Or, nous sommes mis en garde *explicitelement* contre certains de ses défauts :

- le logiciel est en stade de développement beta à propos des AUs sur une vidéo à plusieurs visages (*AUs MultiVid* ).
- lorsque le visage est trop petit sur la vidéo, les repères du visages peuvent être perdus (*Landmark detection*).
- le logiciel ne peut détecter que 4 visages à la fois en mode vidéo à plusieurs visages.<sup>3</sup>

Nous avons par ailleurs remarqué **d'autres défauts** qui compromettent l'utilisation de ce logiciel. OpenFace est parfois capable de **détecter les mains ou les logos comme des visages**. Sur l'exemple figure 2 à gauche, le logo du pull de l'Ecole polytechnique est reconnu comme un visage pendant plusieurs instants temporels (*frames*). Aucun message d'erreur ne nous met en garde. Pour pallier ce problème, nous pouvons vérifier à la main que la détection est correcte. Rappelons aussi que seules certaines *frames* sont pertinentes. Aussi, nous pouvons seulement vérifier les images qui seront exploitées par la suite (i.e aux alentours des réponses Wooclap), voire ne pas considérer les *frames* incorrectes si elles sont en nombre suffisamment limitées.

---

3. Le logiciel peut analyser jusqu'à une vingtaine de visages mais seulement lorsqu'il reçoit une image en entrée et non une vidéo.

Un second problème, lié au précédent, est une **confusion dans les indices des visages**. Sur la figure 2 en haut à droite, l'indice 0 correspond au visage humain et le logo du pull est attribué à l'indice 1 à partir de l'instant 10641, puis quelques secondes après (en bas à droite) le logiciel détecte de nouveau correctement 3 visages. Or, si l'indice 0 correspond au premier visage, l'indice 1 correspond maintenant à un autre visage humain.

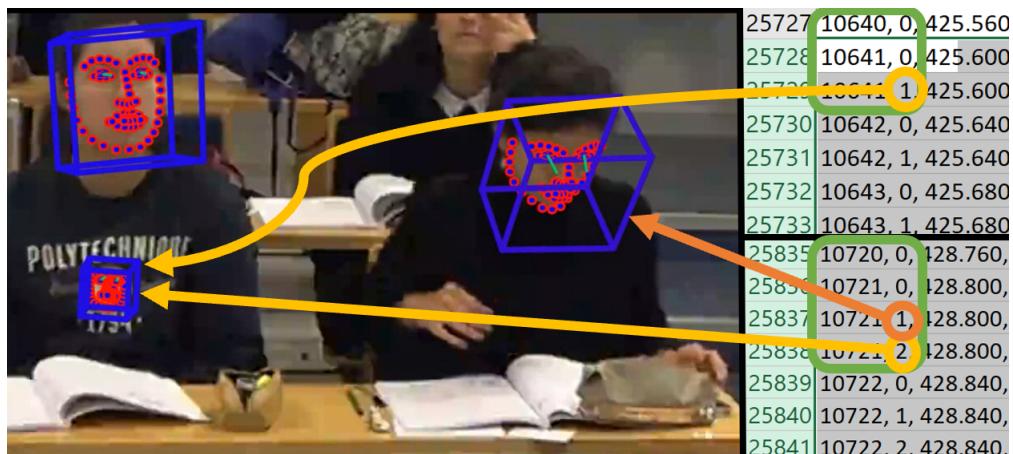


FIGURE 2 – OpenFace reconnaît le logo comme un visage (*gauche*). Le fichier excel correspondant fait alors la confusion entre les indices d'identification (*en haut à droite : lorsque l'id 1 apparaît ; en bas à droite : quelques instants après*).

Ensuite, une fois le fichier excel brut obtenu d'OpenFace, il faut savoir en extraire les données pertinentes. Pour ce faire nous avons du :

- **Sélectionner les seuls critères pertinents** : veut-on mesurer l'agitation des élèves, leur position, la direction de leur regard ? Comment évaluer s'ils sont avachis par exemple, ou s'ils parlent à leur voisin ?

Nous en avons finalement retenu 7 :

1. **Direction du regard** : nous regardons si la ligne de vue de l'élève arrive sur le tableau (par un calcul géométrique simple qui implique de connaître les dimensions de la salle, et le numéro de chaise de l'élève).
2. **Focalisation du regard** : nous calculons le produit scalaire entre les deux vecteurs normalisés "ligne de vue" de chaque œil, fournis par OpenFace. Cela nous permet de savoir si l'élève regarde dans le vide ou non.
3. **Angles de la tête** : comme on peut le voir sur la figure 2, OpenFace enferme chaque visage dans une boîte et calcule ses trois angles de rotation par rapport à la position orthogonale. Cela permet de voir si l'élève est couché sur sa main, penché sur sa table, ou qu'il est tourné sur le côté (ex : pour parler à un voisin).

4. **Agitation** : nous mesurons le déplacement de la tête en rotation et en translation.
5. **Visage neutre** : nous faisons l'hypothèse qu'un visage neutre (le moins possible d'expressions faciales présentes) est associé à un état de concentration.
6. **Sourire** : nous supposons qu'en général, un sourire est plutôt dû à une discussion avec un voisin et donc à une concentration moindre.
7. **Bouche ouverte** : de même, nous faisons l'hypothèse que l'ouverture de la bouche indique une discussion ou un bâillement, et donc une faible attention.

Certains de ces critères ont déjà été étudiés dans la littérature, notamment par Toby Dragon and al. [1], où ils ont été relevés par des expérimentateurs humains présents dans la salle : il sera donc intéressant de comparer nos résultats.

Pour cette raison, nous avons ajouté ultérieurement trois autres critères, *lip corners pull*, *lip pucker* et *eyebrows raise* (que l'on peut traduire par "coins des lèvres retroussés", "lèvres froncées" et "sourcils relevés"), pour tenter de reproduire les résultats de [5], même si ces critères nous semblaient moins intuitifs et se recoupent largement avec les nôtres (en termes d'AU impliqués).

#### — **Créer des outils pour mesurer ces critères :**

Par exemple, il faut déterminer comment quantifier l'agitation. Ici, Nous avons choisi de moyennner les distances euclidiennes parcourues par le centre de la tête de l'élève entre deux images successives.

Une question plus délicate s'est posée pour les **AUs (Action Units)** précédemment évoquées : dans le système FACS (Facial Action Coding System), chaque action musculaire possible est codée par un nombre (ex : AU1 = lever le sourcil gauche). Les différentes expressions faciales ont ensuite été cataloguées et décomposées sur cette base par des experts : chaque mouvement qui participe à l'expression faciale se voit affecté d'une lettre entre A et E traduisant son intensité<sup>4</sup>. Les caractérisations des expressions faciales ainsi formées peuvent diverger un peu d'un auteur à l'autre, mais certaines observations sont relativement consensuelles : le sourire, par exemple, peut être bien traduit par la combinaison 6c-7c-12d-(25b-26b).

Le problème est qu'OpenFace fournit uniquement la présence ou l'absence des AUs, ainsi qu'un nombre décimal représentant leur intensité, mais dont les valeurs typiques varient beaucoup d'un AU à l'autre. La difficulté était alors de construire une mesure pertinente pour chaque expression étudiée à partir de ces données.

#### — **Tester les valeurs de seuils avec des expériences témoins** : à l'issue de l'étape précédente, nous disposons d'une mesure pour chaque variable que nous

---

4. Ce type d'étiquetage a été entrepris depuis les années 70 par de nombreux chercheurs.

souhaitons évaluer. Mais les tests statistiques que nous voulons réaliser par la suite s'appliquent souvent à des variables binaires (présence ou absence d'un caractère), et de plus ces mesures en elles-mêmes sont rarement très explicites : nous n'avons intuitivement aucune idée des valeurs de la focalisation du regard d'un individu, ou de la vitesse moyenne de sa tête lorsqu'il écoute un cours.

Il a donc fallu **établir des seuils de manière arbitraire**, valeurs au-dessus desquelles l'élève est par exemple estimé "agité" ou "regardant dans le vide", tandis qu'en-dessous il est "calme" et "fixe le tableau". Pour ce faire, nous nous sommes filmés nous-mêmes à différents endroits de la classe et avons joué toutes les attitudes étudiées, puis nous avons analysé les valeurs typiques de nos mesures obtenues dans ces différentes situations. Nous avons aussi vérifié que l'ensemble logiciel+mesures+seuils parvenait bien à détecter les différents items étudiés (ex : sourire, bâillement, regards dans le vide) en comparant ses résultats à notre propre traitement manuel.



FIGURE 3 – Extraits de quelques comportements témoins

Les résultats de ce travail d'interprétation des données sont très satisfaisants : même sans surjouer lors de nos tests vidéo, nous avons constaté que nos mesures donnent des valeurs assez stables d'une personne à l'autre, et qui permettent bien de discriminer les différents états. Sur des observables très marquées comme la présence d'un sourire ou l'ouverture de la bouche, nous n'avons obtenu aucun faux positif ou faux négatif sur nos données de test. Par ailleurs, nous avons pu apprécier l'excellent fonctionnement du logiciel OpenFace lorsqu'une seule personne est présente sur la vidéo, excepté dans certaines situations très particulières (ex : si l'élève est complètement penché sur sa table).

Un **second travail** était de créer un programme VBA capable de relier les données OpenFace et Wooclap, en extrayant les premières autour des temps de réponse des élèves et en les traitant comme indiqué précédemment. Une fois cela effectué, nous n'avions en théorie plus qu'à exécuter notre programme sur tous nos fichiers OpenFace.

En pratique, cette tâche s'est avérée bien plus ardue que prévu, car il fallait gérer à la fois la correspondance temporelle entre les deux sources de données et la correspondance entre les différents identifiants des élèves (ID OpenFace, numéro de table et identité réelle), le tout étant entaché de grandes incertitudes.

Il s'est aussi posé un problème d'itération du procédé sur nos données, très fragmentées pour des raisons purement informatiques (volume des fichiers, *crops* nombreux à cause des limitations d'OpenFace, etc). Finalement, une grande partie du travail de nettoyage et d'étiquetage des données a été fait à la main, ce qui en a fait une des étapes les plus chronophages du projet : par exemple, tous les temps de réponse ont dû être réévalués par visionnage des vidéos pour pallier le manque de précision de ceux fournis par Wooclap (renseignés, on le rappelle, par les élèves eux-mêmes, après avoir indiqué leur état, et précis uniquement à une minute près). En ordre de grandeur, un seul crop de 1h30, regroupant 4 élèves au maximum, nécessite une vingtaine d'heures à chacune des machines que nous avons utilisées, et le nettoyage/étiquetage/traitement des données d'un seul cours a demandé plus de 48 heures de travail à 2 personnes.

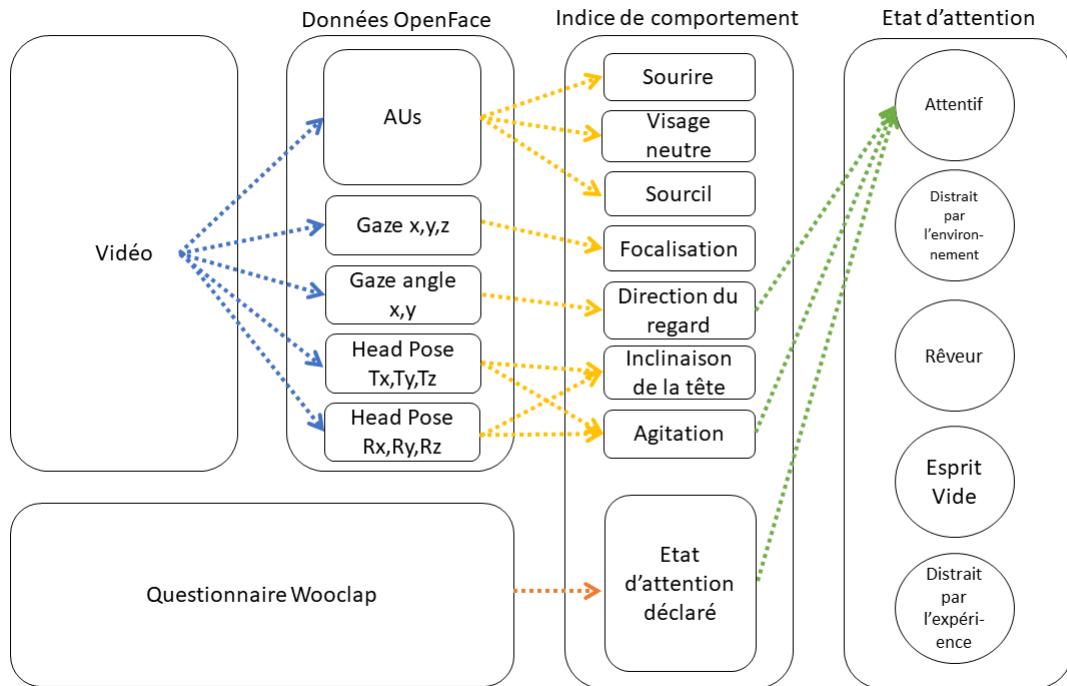


FIGURE 4 – Relations entre les données extraites et les états prédits : traitement par OpenFace (bleu), fonction d'encodage des données OpenFace en indice de comportement (jaune), nettoyage et organisation des données (orange), analyses statistiques pour la prédition (vert).

### 2.3.3 Intégrer une dimension d'étude de terrains auprès des professeurs

Nous souhaitions compléter l'analyse des données de notre expérience en classe par un sondage auprès des professeurs. Les résultats seront analysés indépendamment dans la perspective de rendre plus pertinent les critères et les méthodes employées pendant le projet. (cf. 4)

Du point de vue du projet, cette étude porte la volonté de bénéficier de l'expérience de *terrain* des professeurs. Nous étions aussi intéressés par le souci d'intégrer deux types d'études distinctes : l'expérience en classe et le sondage des professeurs.

## 3 Résultats finaux

### 3.1 Résultats expérimentaux et corrélation

#### 3.1.1 Expérience prototype évaluée

L'expérience prototype menée en classe de CHI 431, sur 2 séances au mois de novembre, avait au moins 3 objectifs :

- Objectif protocolaire : mettre en oeuvre et éprouver un protocole efficace de mesure
- Objectif logistique : se préparer à mener une expérience répétée pendant 10 séances
- Objectif analyse : mener une analyse à la main et évaluer son efficacité

Nous avons de fait rempli les deux premiers objectifs pour mener l'expérience longue en seconde période.

Ensuite, nous avons profité des données vidéos de l'expérience prototype pour essayer une analyse manuelle : faire évaluer l'attention à un observateur extérieur. Il s'agit de discréteriser la vidéo en pas de temps constants (ex : toutes les 2 minutes). A chaque instant, et pour chaque élève présent, un observateur attribue une note de 1 à 10 à chaque élève. Nous avons fait évaluer les 50 premières minutes de vidéo par 3 observateurs indépendants.

#### *De l'efficacité de la méthode à la main*

L'expérience montre d'abord que la méthode est **très longue** (près de 2 heures pour 50 minutes de vidéos). Ensuite, nous avons constaté la difficulté de déceler **seulement par les expressions du visage ou du corps** ce qui se déroule dans la tête de l'élève. De plus, la difficulté à chiffrer un ressenti visuel nous a confirmé l'ambiguité d'une notion floue et abstraite comme l'attention.

#### *Résultats*

La moyenne des écarts-types des notes attribuées à tous les élèves, sur toute la vidéo et entre les 3 évaluateurs est 1,40. Pour une note dans [1, ..., 10], ceci représente une fluctuation entre tous les observateurs assez faible de 14 %. Néanmoins sur un pas de temps, on peut observer une moyenne d'écart-type inférieur à la moyenne totale des écart-types, mais pour certains élèves une variation très grande entre les évaluateurs. Par exemple, à  $t = 00 : 20$  l'écart-type (1.30) est inférieur à l'écart-type moyen (1.40), mais pour l'élève en C7/C19 l'évaluation entre les 3 observateurs donne une incertitude de plus de 2.5 points. L'inconvénient de notre expérience est qu'elle n'a été menée qu'une seule fois et est donc très peu représentative. De plus, 3 évaluateurs ne forment qu'un petit panel. Le panel pourrait être agrandi pour éliminer les valeurs extrêmes et atténuer les variations dans les notes attribuées.

|    | A                               | B     | C    | D    | E    | F    | G           | H | I    | J    | K | L | M                        |
|----|---------------------------------|-------|------|------|------|------|-------------|---|------|------|---|---|--------------------------|
| 1  | Minute :                        | 00:20 |      |      |      |      |             |   |      |      |   |   |                          |
| 2  |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 3  |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 4  | Moyenne sur les 3 évaluateurs : |       |      |      |      |      |             |   |      |      |   |   |                          |
| 5  |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 6  |                                 |       |      | 7,33 | 4,00 | 7,67 |             |   |      |      |   |   |                          |
| 7  |                                 | 8,00  | 6,00 | 9,33 | 9,00 | 7,00 |             |   | 7,00 | 8,67 |   |   |                          |
| 8  |                                 |       |      | 4,00 | 8,33 |      |             |   | 5,33 | 4,33 |   |   |                          |
| 9  |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 10 |                                 |       |      |      |      |      | BLACK BOARD |   |      |      |   |   |                          |
| 11 |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 12 |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 13 |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 14 |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 15 |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 16 | Ecart-type                      |       |      |      |      |      |             |   |      |      |   |   |                          |
| 17 |                                 |       |      |      |      |      |             |   |      |      |   |   |                          |
| 18 |                                 |       |      | 0,58 | 2,00 | 0,58 |             |   |      |      |   |   |                          |
| 19 |                                 | 1,00  | 2,65 | 0,58 | 0,00 | 2,00 |             |   | 1,00 | 0,58 |   |   | Moyenne Ecart-Type       |
| 20 |                                 |       |      | 2,00 | 0,58 |      |             |   | 2,08 | 2,52 |   |   | 1,30                     |
| 21 |                                 |       |      |      |      |      | BLACK BOARD |   |      |      |   |   |                          |
| 22 |                                 |       |      |      |      |      |             |   |      |      |   |   | Moyenne Ecart-Type Total |
| 23 |                                 |       |      |      |      |      |             |   |      |      |   |   | 1,40                     |

FIGURE 5 – Exemple des variations d'incertitudes à  $t = 00 : 20$ 

### Conclusion

La méthode est assez lourde et longue. La distribution des notes attribuées est en moyenne assez resserrée. Mais la fiabilité de telles évaluations individuelles extérieures est relative. Ensuite, l'expérience est peu reproductible. Que se passerait-il en effet si nous prenions 3 nouveaux évaluateurs ? Pour conclure, cette méthode représente toujours une solution correcte et exploitable pour évaluer l'attention des élèves. Mais évaluer à la main n'est pas viable pour un trop grand jeu de données ( $\geq 3h$  de vidéos). Enfin, évaluer depuis l'extérieur est toujours difficile et dans la perspective à long terme d'avoir un outil automatisé, nous nous sommes concentrés sur le développement d'une méthode automatique par ordinateur.

### 3.1.2 Expérience longue

#### *De la qualité des résultats obtenus*

L'ambition était de suivre une période de cours entière (10 séances initialement). Or, plusieurs facteurs externes ont nui à la quantité et la qualité des réponses récoltées : cours annulés, événements de la vie associative, manque d'assiduité et manque de coopération des fournisseurs du matériel (fig. 6).

| Date                | 7/12   | 14/12  | 20/12  | 11/01  | 18/01  | 25/01  | 1/02   | 8/02   | 15/02  |
|---------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Cours               | annulé | assuré | annulé |
| Matériel disponible |        | oui    | oui    | oui    | oui    | oui    | oui    | non    |        |
| Réponses assidues   |        | oui    | oui    | non    | oui    | oui    | non    |        |        |

FIGURE 6 – Tableau résumant la réalisation des expériences. Vert : Expérience possible. Rouge : Expérience annulée. Noir : Expérience arrêtée

#### *Du croisement des réponses au sondage et des présents*

Nous avons aussi pâti d'un défaut de correspondance entre les numéros des places renseignés par les élèves et les places réellement occupées dans la salle. De notre côté, l'organisation de l'expérience aurait dû nécessiter davantage de vigilance sur ce point. Par exemple, en vérifiant systématiquement auprès des élèves pendant l'expérience le bon remplissage des informations. Les numéros des places auraient dû être fixés sur les tables pendant toute la période. Au contraire, rappeler simplement à l'oral l'indexation des places augmente le risque de confusion (Fig. 7).

Ce problème peut néanmoins être contourné en identifiant les élèves à la main à partir de leurs temps de réponse observés sur la vidéo : c'est ce que nous avons entrepris, mais le procédé est assez chronophage.

Au final, nous disposons de données en quantité suffisante pour les séances des 14/12, 21/12, 18/01 et 25/01. De plus, en raison du temps nécessaire à l'analyse complète : **tous les traitements statistiques impliquant des données OpenFace se concentrent sur les données du 21 décembre.**

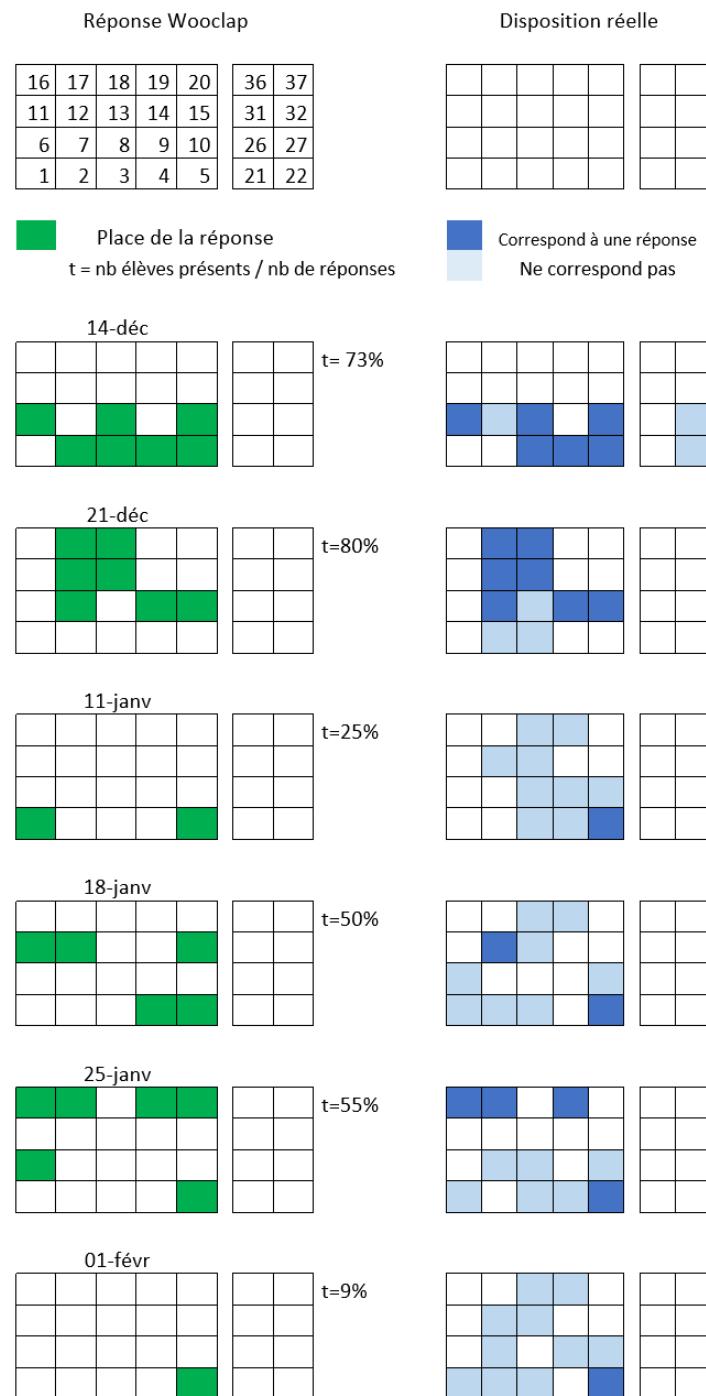


FIGURE 7 – Qualité des réponses Wooclap et correspondance spatiale

## 3.2 Modèle et analyse

### 3.2.1 Analyse des réponses au sondage seules

#### *Cadre théorique*

Cette partie vise à utiliser des outils de mathématiques appliquées rigoureux pour formaliser l'intuition suivante : **les changements d'états mentaux les plus fréquents sont les plus probables.** On peut donc s'appuyer sur l'analyse des fréquences pour en déduire la probabilité d'être dans chaque état. Nous avons choisi un modèle simple, s'inspirant de celui d'Ising, qui offre un cadre théorique pour décrire les transitions de phase d'attention. Le cadre mathématique exact est développé en annexe A.

Les résultats du sondage forment un ensemble d'états qui sont des réalisations d'une chaîne de Markov. Dans une première approximation, on suppose que tous les élèves changent d'état de façons identiques. On peut ainsi calculer  $\mu^o(x; n)_i$ , la fréquence de l'état  $x$  pour un élève  $e_i$  et moyenner sur tous les élèves. On obtient le tableau de valeurs suivantes pour les 209 réponses :

| Etat $x$      | Attentif | Distrait par l'environnement | Rêveur | Esprit vide | Distrait par l'expérience |
|---------------|----------|------------------------------|--------|-------------|---------------------------|
| $\mu^o(x; n)$ | 0.65     | 0.19                         | 0.09   | 0.02        | 0.05                      |

Tableau de la mesure de probabilité invariante sur F

#### *Probabilités de transition entre les états*

Nous avons cherché à mettre en évidence l'existence de transitions privilégiées.

Une première intuition consiste à relever **la fréquence de chaque transition possible.**

$$\text{Pour } x, y \in F, \text{ on pose } P^o(x, y; n) = \frac{\text{Card}\{0 \leq l \leq n-1 : X_l = x, X_{l+1} = y\}}{\text{Card}\{1 \leq l \leq n : X_l = x\}}$$

Nous n'avons considéré que des transitions sur des intervalles de 5 à 15 minutes. En effet, certains élèves renseignaient à la fin d'une séance jusqu'à 4 états en moins de 5 minutes (c'était souvent 4 fois le même). On a estimé que ce comportement était différent du comportement sur des intervalles supérieurs à 5 minutes. Une fois les calculs effectués, on obtient une matrice de transition (fig.8).

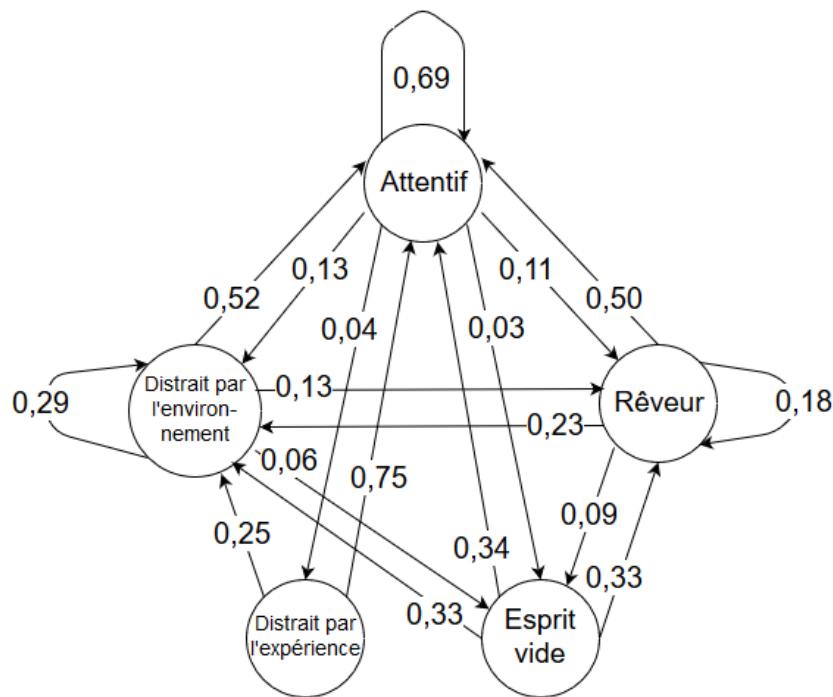


FIGURE 8 – Graphe de transition

### Résultats

On remarque tout d'abord que certains états sont peu ressentis par les élèves. En particulier, l'état "Esprit vide" n'a été sélectionné que 2 % des instants. Les élèves ne se déclarent distraits par l'expérience seulement 5% des instants. Ce résultat bienvenu qui semble minimiser le biais de l'expérience est à relativiser. En effet, les élèves sondés après l'ensemble des expériences ont **majoritairement reconnu que l'expérience était "intrusive"** ou les "dérange[ait]". A l'inverse, dans cette théorie, nous affirmons que la mesure invariante de l'état "Attentif" est égale à 0,65. **Les élèves ont tendance à être attentifs.** A propos des transitions, le pourcentage peut cacher la fréquence absolue d'une transition. Par exemple une probabilité de passage de 0,75 n'est pas significative si elle correspond à 3 transitions sur 4, alors qu'une de 0,5 qui correspond à 72 sur 144 est plus significative. Ainsi, la faible quantité de données et la faible qualité des réponses engagent la fiabilité d'une telle exploitation. En effet, davantage de réponses auraient permis d'exclure plus sûrement certaines transitions comme des fluctuations rares. Nous mentionnons ici celles réalisées très rarement (sans considérer l'état 4) : 2 → 5, 3 → 5, 5 → 2, 5 → 5.<sup>5</sup>

Toutefois, nous avons mis en évidence certains chemins préférentiels : 2 → 1, 3 → 1,

5. {1 = "attentif", 2 = "distract par l'environnement", 3 = "rêveur", 4 = "esprit vide", 5 = "distract par l'expérience"}

$5 \rightarrow 1, 3 \rightarrow 2.$

La probabilité d'être attentif à la fin de l'intervalle en ayant été attentif au début est de 0.69. On retrouve une valeur similaire aux résultats de Mirko Raka (0.58) [3].

#### *Vers une autre méthode statistique : Modèle Markov caché*

Dans la lignée de ce travail sur les chaînes de Markov, nous nous étions tournés vers les *chaînes de Markov cachées* (ou MMC pour Modèle de Markov caché), qui nous paraissaient particulièrement adaptés à notre expérience. Dans un MMC, on suppose qu'une suite de variables d'intérêt ( $X_n$ ) (ici, la suite des réponses Wooclap d'un élève) suit un modèle de Markov inconnu, et que l'on dispose d'une suite de variables expérimentales observables ( $Y_n$ ) (ici, les résultats OpenFace de l'élève). On suppose que cette deuxième suite suit un certain type de loi (par exemple une gaussienne), dont les paramètres sont également inconnus, et que ses termes sont identiquement distribués. A partir des valeurs observées de  $X_n$  et  $Y_n$ , un algorithme assez complexe permet de calculer la matrice de transition de ( $X_n$ ) et les paramètres des lois  $Y_n$ .

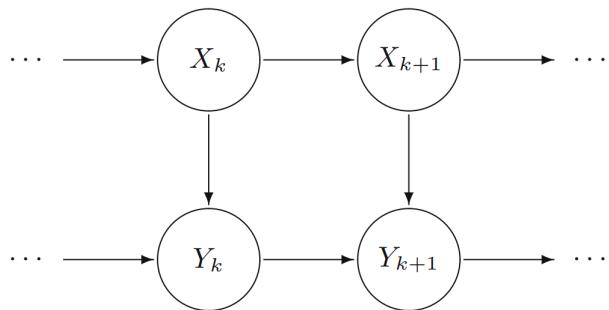


FIGURE 9 – Illustration du modèle Markov Caché

Ce traitement aurait donc été très intéressant pour nous, car il nous aurait donné des informations sur le comportement de l'attention des élèves *en tenant compte des variables OpenFace*, sans avoir à tester préalablement la pertinence de ces variables par une analyse de corrélations (les paramètres de convergence nous auraient même renseignés *a posteriori* sur cette pertinence). Cependant, avec comme priorité une obtention de résultats fiables sur le comportement de l'individu, et sous les conseils de M. Bodineau, professeur de modélisation des phénomènes aléatoires, utiliser le MMC à ce stade-là de la réflexion était une mauvaise option. Ce serait, en d'autres mots, spéculer sur un comportement très précis (bien que potentiellement correct), et chercher à adapter un algorithme relativement compliqué qui nécessite une quantité de données inatteignable en si peu de temps. Nous

gardons donc cette idée comme développement possible de notre projet, mais n'avons pas cherché à l'implémenter.

### ***Evolution dans le temps***

Nous avons cherché à mettre en évidence l'évolution temporelle du niveau d'attention. Nous simplifions les réponses en ne considérant que les élèves se déclarant "attentif" parmi les cinq états. Il est à noter que les élèves sont libres de répondre quand ils veulent avec la consigne de répondre toutes les 12 minutes grâce à une alarme sur leur téléphone. Mais cette alarme est parfois mal programmée, non remarquée ou simplement ignorée. Les réponses ne sont donc pas distribuées uniformément sur tous les intervalles. Nous choisissons donc de représenter les histogrammes avec le nombre de réponses "attentif" comparé au nombre de réponses total sur un intervalle de 10 minutes.

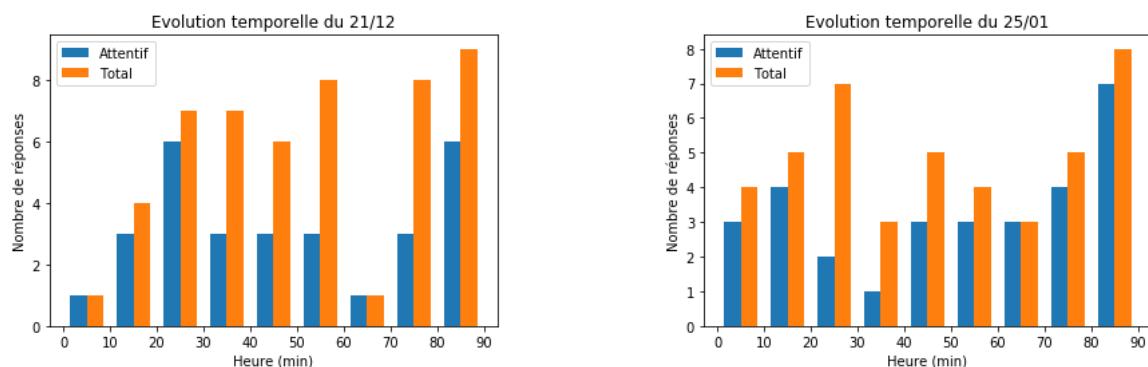


FIGURE 10 – Histogramme d'évolution temporelle. Les deux séances du 21/12 (*gauche*) et 25/01 (*droite*).

Les deux séances du 21/12 et du 25/01 semblent mettre en évidence **une évolution similaire** : un pic d'attention dans la première demi-heure, une décroissance jusqu'au 60 premières minutes, puis dans les dernières 15 minutes, la proportion d'élèves attentifs augmente de nouveau.

Les autres séances ne présentent pas de profil temporel aussi franc et sont inexploitables à cause de la disparité des mesures.

### ***Matrice dans l'espace***

Nous nous sommes aussi intéressés à l'influence qu'avait l'environnement des élèves sur les transitions d'un état à l'autre. Les résultats obtenus lors du cours de CHI 421 étaient difficilement exploitables car les élèves ne répondaient pas tous aux mêmes instants, ce

qui ne nous permettait pas de connaître facilement l'état de toute la classe au moment de chaque réponse. Pour savoir si se lancer dans cette analyse serait pertinent, nous avons donc d'abord étudié cette influence pour les cours d'HSS 414b des 28 novembre et 13 décembre 2018.

#### *Données collectées*

Le 28 novembre, il y avait 16 élèves et nous avons recueilli respectivement 12, 10, 7 et 6 réponses à 8h30, 9h00, 9h30 et 10h00 ; le 13 décembre, il y avait 17 élèves et nous avons cette fois recueilli respectivement 17, 16, 16 et 13 réponses. Si ces résultats sont très critiquables - faible nombre de réponses, forte tendance dans l'absolu à répondre "attentif" (76% puis 69% des réponses) et durée importante entre les réponses -, ceux-ci nous permettent d'évaluer si l'attention globale de la classe a eu un effet sur l'évolution de l'attention individuelle de chaque élève.

#### *Méthode*

Les données collectées lors de ces expériences ne permettaient pas de situer spatialement les élèves dans la classe et ainsi de pondérer les réponses par proximité : donc nous avons évalué l'impact de l'attention globale de la classe sur celle d'un élève  $e_i$  donné.

Pour un élève donné, pour une transition d'état observée, nous calculons le pourcentage des autres élèves dans la classe qui se déclarent attentifs. La matrice représente le résultat en fonction de la nature des transitions. Le signe '-' indique que la transition n'a pas été observée. La matrice b figure 9 se lit ainsi pour la quatrième ligne et la cinquième colonne : en moyenne, pour la transition d'un élève de l'état  $E_t$  "rêveur" à l'instant t, vers l'état  $E_{t+30}$  "distrait par l'expérience" à l'instant t+30, 80% des autres élèves de la classe se déclaraient attentifs à l'instant t initial.

a.

| $E_t$                        | $E_{t+30}$ | Attentif | Distrait par l'environnement | Rêveur |
|------------------------------|------------|----------|------------------------------|--------|
| Attentif                     |            | 80%      | 91%                          | 86%    |
| Distrait par l'environnement |            | 73%      | -                            | -      |
| Rêveur                       |            | 89%      | -                            | -      |

b.

| $E_t$                        | $E_{t+30}$ | Attentif | Distrait par l'environnement | Rêveur | Distrait par l'exp. | Esprit vide |
|------------------------------|------------|----------|------------------------------|--------|---------------------|-------------|
| Attentif                     |            | 70%      | 72%                          | 80%    | -                   | -           |
| Distrait par l'environnement |            | 80%      | -                            | -      | -                   | -           |
| Rêveur                       |            | 63%      | -                            | 76%    | 80%                 | -           |
| Distrait par l'expérience    |            | 63%      | -                            | -      | 87%                 | -           |
| Esprit vide                  |            | 80%      | -                            | 63%    | -                   | 63%         |

c.

| $E_t$                        | $E_{t+30}$ | Attentif | Distrait par l'environnement | Rêveur |
|------------------------------|------------|----------|------------------------------|--------|
| Attentif                     |            | +4       | +15                          | +10    |
| Distrait par l'environnement |            | -3       | -                            | -      |
| Rêveur                       |            | +13      | -                            | -      |

d.

| $E_t$                        | $E_{t+30}$ | Attentif | Distrait par l'environnement | Rêveur | Distrait par l'exp. | Esprit vide |
|------------------------------|------------|----------|------------------------------|--------|---------------------|-------------|
| Attentif                     |            | +1       | +3                           | +11    | -                   | -           |
| Distrait par l'environnement |            | +11      | -                            | -      | -                   | -           |
| Rêveur                       |            | -6       | -                            | +7     | +11                 | -           |
| Distrait par l'expérience    |            | -6       | -                            | -      | +18                 | -           |
| Esprit vide                  |            | +11      | -                            | -6     | -                   | -6          |

a. Pourcentage moyen d'élèves attentifs à la date  $t$  autour d'un élève dans l'état  $E_t$  à la date  $t$  passant à l'état  $E_{t+30}$  à la date  $t+30$  le 28 novembre. b. *Idem* le 13 décembre.

c. Points d'écart au pourcentage moyen d'élèves attentifs à la date  $t$  autour d'un élève dans l'état  $E_t$  à la date  $t$  passant à l'état  $E_{t+30}$  à la date  $t+30$  le 28 novembre. d. *Idem* le 13 décembre.

FIGURE 11 – Influence de l'attention collective en HSS414b

### Premières remarques sur les matrices

Pour la figure a, le 28 novembre, l'état "Distrait par l'expérience" n'a été choisi qu'une fois et en début d'expérience, et l'état "Esprit vide" n'a jamais été choisi, ce qui explique pourquoi ces états n'apparaissent pas dans le tableau. En ce qui concerne les figures c et d, le 28 novembre, le pourcentage moyen d'élèves attentifs était de 76% et une différence d'un élève se traduisait par un écart pouvant varier de 9 à 20 points, tandis que le 13 décembre, le pourcentage moyen d'élèves attentifs était de 69% et une différence d'un élève se traduisait par un écart pouvant varier de 6 à 9 points.

### Résultats

Ces résultats ne semblent donner **aucune tendance significative**. Parfois, la valeur du pourcentage est même contre intuitive, comme figure a, lorsque les élèves passent de "attentif" à "distrait par l'environnement" alors qu'à l'origine 91% de la classe était attentive.

Une telle contradiction provient probablement de nouveau de la qualité trop faible des données et de la méthode de mesure : en 30 minutes, de multiples transitions d'états sont certainement possibles et nous ne les mesurons pas.

Nous n'avons alors pas cherché à exploiter les données recueillies lors des cours de CHI 421 de l'expérience longue. Néanmoins, cela n'invaliderait en rien la démarche que nous avons menée, et il pourrait être intéressant l'an prochain de la reproduire avec un échantillon de réponses plus massif et dense. Si cela est même possible, coupler cette méthode à une pondération par proximité des réponses obtenues donnerait des résultats encore plus riches : on pourrait ainsi comparer l'effet de l'environnement immédiat et celui de l'environnement global.

### ***Conclusion générale***

Les données sont peu nombreuses et de trop faible qualité pour conclure. Par exemple, le nombre de mesures (209) est faible pour obtenir une moyenne significative. Rapelons aussi que cette valeur cache l'hypothèse d'avoir confondu tous les élèves et toutes les expériences entre elles. Or, comme le montre la figure 7, les séances ne sont pas équivalentes entre elles du point de vue de la disponibilité des élèves. Ensuite, il convient de s'interroger sur le sens à donner à des transitions sur des intervalles de 15 minutes. Evidemment ce criblage n'est pas assez élevé pour espérer détecter toutes les transitions d'états. Malheureusement, un criblage temporel plus fin est irréalisable à cause du caractère intrusif du dispositif. Enfin, un sondage post-expérience auprès des élèves a montré que tous ont jugé l'utilisation régulière du téléphone comme troublante.

### **3.2.2 Analyse couplée des réponses et des vidéos**

#### ***Relevé manuel***

Nous relevons manuellement par un observateur les types d'action effectués par les élèves dans la minute précédant leur réponse Wooclap. Nous avons choisi les types d'actions en accord avec les indices corporels que la méthode par ordinateur est capable de mesurer et en inspiration avec des travaux similaires par Mirko Raca [4]. Nous avons sélectionné les types d'actions suivants :

- Concentré cours : Ecoute le cours ou est concentré sur le tableau
- Prise de notes : Ecrit sur son cahier
- Figure prostrée : Tête sur la main, corps avachi, signes de sommeil
- Bavardage : Discute avec ses camarades
- Concentré autre : Concentré sur un autre medium

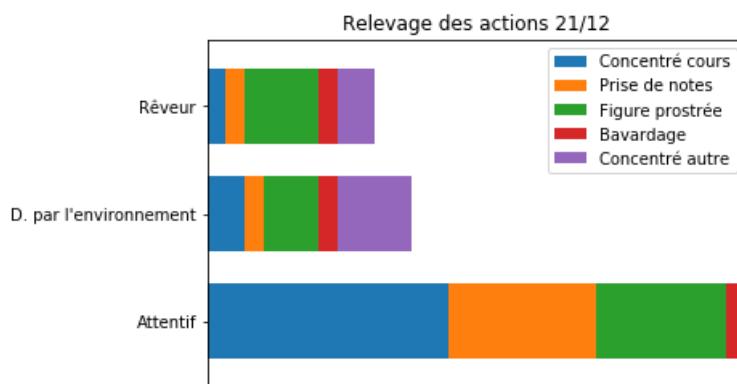


FIGURE 12 – Distribution des actions en fonction des états le 21/12

### *Résultats*

Les états peu utilisés (tête vide et distractif par l’expérience) sont négligés. Les barres sont normalisées entre elles de façon à ce que leur somme égale 1, mais la représentation contient volontairement encore l’écart du nombre de données entre l’état ”attentif” et les autres. Les élèves se disent plus souvent ”attentifs” et notre jeu de données ne contient que très peu de réponses pour les autres états.

L’histogramme confirme l’intuition que les élèves attentifs sont majoritairement en train d’écouter le cours ou de prendre des notes. Néanmoins, certains se déclarent attentifs alors qu’ils ont leut tête posée sur leur main (‘figure prostrée’). De plus, l’histogramme suggère que les états ’Rêveur’ et ’Distrait par l’environnement’ sont peu distinguables du point de vue de la distribution des actions.

L’observateur peut être biaisé lors des évaluations car il est difficile d’attribuer une action aussi vague, sur un intervalle aussi large et mal circonscrit, à un élève. Si l’élève se déclare plutôt ’rêveur’, l’évaluateur peut avoir tendance à décrire la personne comme concentré sur un autre support plutôt que le cours sans de réel indice visuel.

### *Par outils d’analyse statistique*

L’objectif ici est d’étudier les corrélations statistiques entre nos deux jeux de données, d’une part les réponses Wooclap et d’autre part les données extraites des fichiers Open-Face comme expliqué en section 2. Nous nous sommes naturellement tournés vers les corrélations ”classiques” de Pearson, mais pour des raisons que nous expliquerons plus loin, celles-ci sont peu adaptées à notre protocole, ce qui nous a motivés à appliquer d’autres méthodes. Avant de pouvoir appliquer ses méthodes, nous dressons en revue le cheminement parcouru et les choix effectués.

Deux points valent d'être précisés : (1) on rappelle que les données OpenFace que nous utilisons sont extraites **une minute avant la réponse de l'élève**, pour éviter d'analyser les attributs physiques de l'élève au moment où il est penché sur sa table et écrit sur son portable ; (2) les données sont moyennées sur environ 30 secondes, pour éliminer les artefacts dus à OpenFace et les gestes non-significatifs (par exemple, un très bref déplacement qui ferait conclure à un élève agité).

#### *Robustesse et validité des variables explicatives*

Dans notre démarche, avoir une quinzaine de variables explicatives rend difficile la tâche de trouver des relations *a priori* inconnues. Tester la pertinence de ces variables sans déterminer la loi sous-jacente serait un gain, car il validerait le choix et orienterait le travail vers une recherche de lois plutôt que vers une recherche de nouvelles variables.

Il se trouve qu'une approche de classification non-supervisée permet cette assertion. En entrée, l'algorithme récupère à la fois les données d'entrée de variables Openface et la donnée Wooclap, et renvoie le "graphe de communautés" conséquent, outil de visualisation permettant d'exhiber les corrélations existantes.

Le graphe de communautés choisi ici, est un graphe dont l'ensemble des sommets représente une donnée indépendante (Réponse Wooclap et sortie Openface). Une arête entre deux sommets est choisie selon la similarité entre les données OpenFace, tandis que la couleur du sommet est déterminée par la réponse Wooclap (vert si attentif et rouge sinon). Le graphe engendré se compose naturellement de deux communautés caractérisées par la couleur des sommets. Au sein d'une communauté, plus le nombre d'arêtes est dense, plus la communauté est cohérente.

Un graphe avec deux blocs denses distincts de sommets de couleur unie se traduit donc par deux classes de données où, dans chaque classe, les réponses Wooclap corrèlent fortement avec les sorties OpenFace.

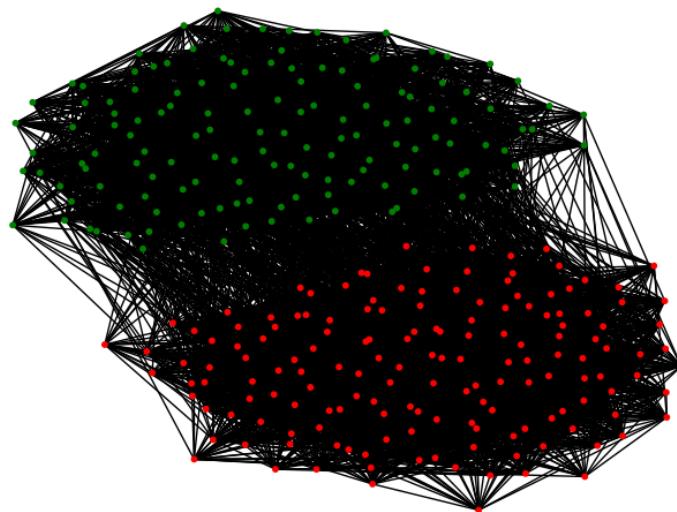


FIGURE 13 – Classification non-supervisée avec toutes les variables explicatives. On commence à distinguer deux communautés de données, bien que la distinction entre les deux blocs ne soit pas flagrante. La conclusion immédiate est que nous pouvons affiner notre recherche des variables pertinentes parmi ce lot initial de quinze variables.

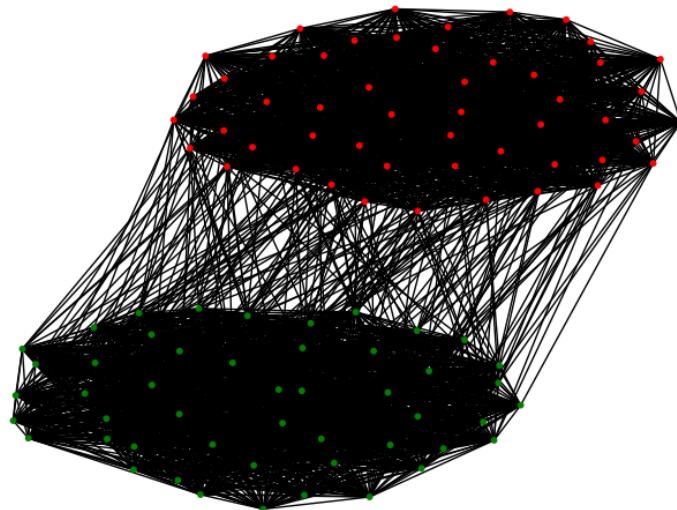


FIGURE 14 – Classification non-supervisée avec les 7 critères retenus (cf. 2.3.2). Les communautés se distinguent clairement ici. Ceci implique que la restriction à ces quelques variables se révèle pertinente.

### *Par analyse des quasi-implications*

Dans le cadre de nos expériences, les corrélations classiques souffrent d'un défaut majeur : elles marchent "dans les deux sens", c'est-à-dire qu'elles donnent un résultat significatif seulement lorsque les deux traits étudiés sont toujours (ou presque) présents simultanément chez l'individu.

Mais nous cherchons ici à mettre en évidence des phénomènes plus fins que cela : par exemple, nous voudrions savoir si un élève est attentif lorsqu'il regarde le tableau, ou encore mieux, lorsqu'il regarde le tableau et que son regard est focalisé. Supposons que l'élève, lorsqu'il est attentif, regarde le tableau la moitié du temps et son polycopié l'autre moitié : le coefficient de Pearson entre "attention" et "regard tableau" vaudra alors 0 et ne nous dira rien de la relation sous-jacente entre ces deux variables. Quant à étudier le lien entre trois variables comme l'attention, la direction de regard et la focalisation, cet outil ne le permet tout simplement pas.

C'est en fait une structure d'implication que nous cherchons à mettre en évidence : "Regarde le tableau & A le regard focalisé  $\Rightarrow$  Attentif". Nous avons donc cherché des outils statistiques capables de déceler de telles corrélations unilatérales entre variables **binaires**, et nous nous sommes appuyés sur des travaux portant sur les quasi-implications et la logique floue ([2]).

### *Leur principe*

L'idée est de décomposer l'implication à tester en une **conjonction d'implications logiques élémentaires**. C'est la véracité de ces implications élémentaires qui est calculée de manière "floue" : pour chacune d'elles, de la forme  $p \Rightarrow \emptyset$ <sup>6</sup>, on calcule un certain **indice** à partir des fréquences de nos variables observées dans notre échantillon. Plus précisément, pour tout **profil**  $p = ijk\dots$ , avec  $i, j, k, \dots \in \{Vrai, Faux\}$ , où  $i, j, k\dots$  sont nos observables et  $p$  un profil obtenu en précisant la valeur de tous les observables (présent ou absent), on calcule :

$d_{p \Rightarrow \emptyset} = 1 - \frac{f_p}{f_i f_j f_k \dots}$ , où  $f_p$  est la fréquence du profil  $p$  et  $f_i, f_j, f_k \dots$  les fréquences marginales associées aux événements  $i = Vrai$ , etc, observées dans l'échantillon.

Si cet indice est supérieur à une certaine valeur seuil positive fixée au préalable (généralement 0,5), on considère que la proposition  $p \Rightarrow \emptyset$  est quasi-vraie ; sinon, elle est quasi-fausse. En faisant ce calcul sur *tous* les profils  $p$  possibles, on remonte jusqu'à la proposition initiale pour savoir si elle est quasi-vraie ou quasi-fausse. On peut montrer qu'on obtient alors une structure logique cohérente (elle respecte notamment la transitivité).

Supposons que nous étudions seulement deux variables,  $a$  et  $b$  : on veut montrer que  $a$  quasi-implique  $b$ .  $a \Rightarrow b$ , qui s'écrit aussi  $a \neg b \Rightarrow \emptyset$ , est déjà une implication élémentaire. On constate que  $d_{a \neg b \Rightarrow \emptyset}$  vaut 1 si on a  $a \Rightarrow b$  en logique classique (profil  $a \neg b$  absent) , 0 si

---

6. L'ensemble vide désigne l'élément faux.

a et b sont indépendantes (i.e décorrélées), et qu'elle est positive si le profil  $a-b$  est sous-représenté, ce qui va dans le sens de notre hypothèse (et négative s'il est sur-représenté, ce qui y est contraire). Le passage à la logique floue constitue donc un assouplissement de la logique classique : pour déterminer si  $a \Rightarrow b$ , au lieu de se demander si le profil  $a-b$  est absent de nos données, on observe s'il y est *sous-représenté au-delà d'un certain seuil*.

#### *Appliquée à nos données*

Ce test logique est relativement simple en termes de calcul, et permet exactement les manipulations que nous souhaitons. Le seul obstacle est que le temps de calcul croît rapidement avec le nombre  $n$  d'observables, puisqu'il faut effectuer un calcul sur  $2^n$  profils. Nous avions au final  $n = 8$  variables exploitables, ce qui reste très gérable.

En revanche, nous nous sommes heurtés à la faible quantité de nos données : nous ne pouvons pas calculer la fréquence de 256 profils en disposant de moins d'observation que cela. Mathématiquement parlant, cela ne pose pas de problème<sup>7</sup> (nous aurons seulement davantage de profils absents), mais les résultats s'en trouveront biaisés (nous observerons des corrélations artificiellement fortes du fait de cette absence). Nous pourrons donc discuter les valeurs de seuil choisies.

La méthode a été implémentée et reste en cours d'amélioration. Nous sommes face à une difficulté liée au très grand nombre de profils absents. Ceci écrase certains calculs de fréquences et rend imprécis les métriques, ce qui donne des résultats grossiers. Nous cherchons actuellement à y remédier en regardant plus précisément la localisation des fréquences pertinentes.

Néanmoins, nous avons déjà réussi ainsi à retrouver le résultat :

**"visage neutre  $\Rightarrow$  attentif "** : True

#### *Analyse de variance (ANOVA)*

Il s'agit d'une autre méthode d'analyse statistique qui permet d'exhiber des corrélations entre des variables qualitatives, telles que les réponses au questionnaire Wooclap, et des variables quantitatives, telles que les critères OpenFace. Nous avons employé cet outil après avoir étudié les différentes possibilités d'études statistiques (corrélation de Pearson, rho de Spearman, tau de Kendall). Ces dernières peuvent être utilisées pour des variables qualitatives mais n'étaient pas efficaces à cause du petit nombre d'états possibles.

Grâce à l'ANOVA, nous pouvons déterminer si les différents états Wooclap influencent les données OpenFace. Nous avons utilisé le langage de statistiques R qui permet de coder cette analyse très facilement. Nous regroupons tout d'abord les données OpenFace en fonction des **facteurs explicatifs**, qui sont ici les 5 états. Nous calculons ensuite la moyenne du critère au sein de chaque groupe, ainsi que sa variance.

---

7. A condition d'avoir des fréquences marginales non nulles

Ensuite, nous cherchons à savoir à quel point la variation au sein des groupes est plus ou moins importante que celle entre les groupes. Si elle l'est beaucoup moins, il faut alors rejeter l'hypothèse selon laquelle les facteurs explicatifs n'influencent pas le critère. Dans notre cas, on saura alors qu'un certain critère OpenFace est corrélé à l'état d'attention de l'élève.

Cependant on ne sait pas encore quel état Wooclap est différent des autres, il faut alors mener un test **post hoc** pour calculer les différences de moyennes deux à deux entre les différents facteurs.

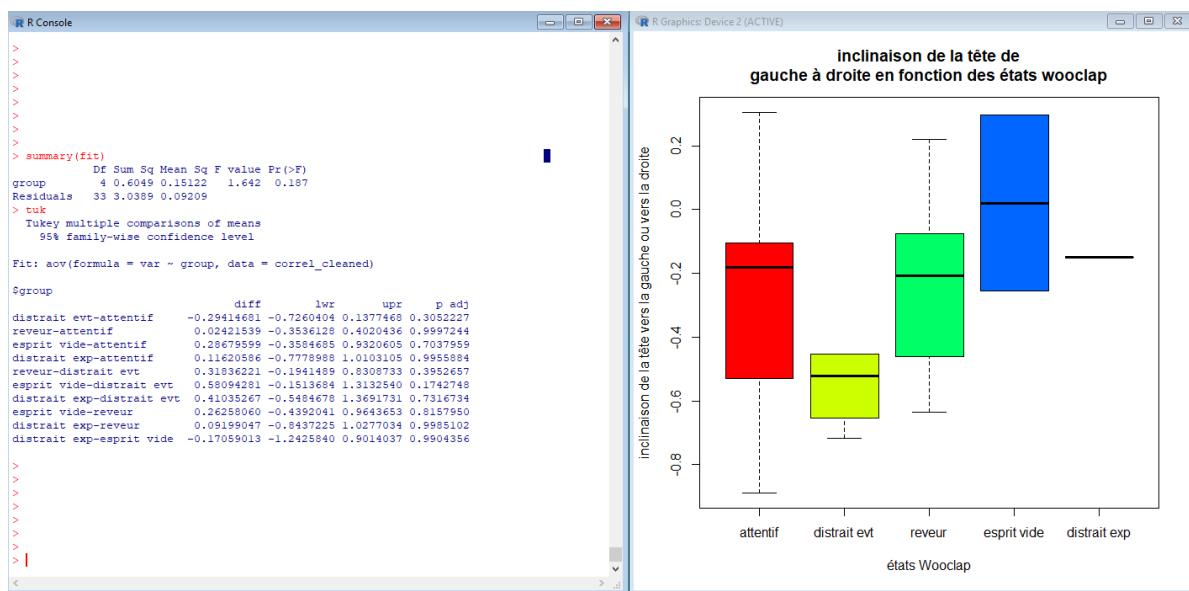


FIGURE 15 – Résultats de l'analyse de variance à propos de l'inclinaison de la tête. À gauche la console R avec entre autres la p-value et les tests post hoc, à droite une représentation de l'inclinaison en fonction des états Wooclap, la moyenne est la barre noire

On remarque que nos marges d'erreurs sont importantes (conséquence du petit nombre de données) et qu'il y a donc très peu de p-value significatives.

Nous avons cependant obtenu une p-value de 0,053 à propos de la corrélation entre le visage neutre et l'attention, ce qui corrobore les résultats de la littérature et de la proposition 'visage neutre  $\Rightarrow$  attentif' : True.

## 4 Croisement au témoignage empirique des professeurs

### 4.1 L'intuition d'un savoir-faire ou plutôt d'un *savoir-détecenter*

Nous partageons l'intuition que les professeurs possèdent déjà un savoir-faire, ou plutôt un *savoir-détecenter*, pour évaluer dans quelle mesure une classe est attentive. En parallèle de nos expériences en classe, nous avons donc également proposé un sondage à destination des professeurs pour connaître leurs « sentiments, intuitions et expériences concernant les modalités d'attention des élèves qu'[ils côtoient] ou [ont] côtoyés au cours de [leur] carrière ». Les réponses aux trente-et-une questions ont été analysées et, en retour, nous ont permis de concevoir les questionnaires destinés aux élèves et d'avoir un regard critique sur les paramètres que nous étudions avec les logiciels de traitement d'images.

### 4.2 Elaboration du sondage

Nous avons composé un sondage en trois parties :

- un **préambule** dans lequel les professeurs peuvent saisir leurs données personnelles (nom, prénom, courriel) et indiquer leurs attentes vis-à-vis du sondage (aucune attente, échanger avec nous, recevoir les résultats...) ;
- une **partie pratique** dans laquelle ils renseignent leur nombre d'années d'expérience, l'âge et le nombre de leurs élèves et répondent à quatre questions ;
- une **partie subjective** sur les critères qu'ils jugent pertinents pour évaluer l'attention, les jours et les moments les plus propices pour faire cours.

#### *Regard critique*

Les trente-et-une questions proposées - nous le voyons aujourd'hui - ne sont pas nécessairement toutes pertinentes, ni le format de réponses que nous avons proposé optimal. Certaines données n'ont pas été exploitées, d'autres seulement partiellement, mais nous savions dès le départ qu'il en serait ainsi. Ou plutôt, la précocité de la conception du sondage par rapport à l'avancement du projet<sup>8</sup> nécessitait que nous établissions un panel de questions suffisamment large pour ne pas avoir besoin de le modifier au cours du projet. Par exemple, dans cette étude, tous les cours que nous avons étudiés et filmés se déroulent à des horaires fixes. Ainsi, les questions sur les jours et horaires les plus propices pour faire cours n'ont plus d'intérêt pour l'étude. Néanmoins, en soi, celles-ci nous apprennent que

---

8. Proposé aux académies à partir de fin novembre 2018

les professeurs préfèrent travailler le mardi<sup>9</sup> ou que choisir des plages de deux heures leur semble trop long.

### 4.3 Résultats

Distribué auprès de professeurs des académies d'Amiens et Bordeaux entre le 10 janvier et le 31 mars 2019, ce sondage a recueilli 1 045 réponses. Les professeurs sondés avaient en moyenne une quinzaine d'années d'expérience<sup>10</sup> - que ce soit auprès d'enfants, d'adolescents, de jeunes adultes ou de détenus - et ont répondu avec enthousiasme à nos questions.

Pensez-vous que l'âge des élèves soit déterminant dans leur capacité d'attention ?



Pensez-vous que le nombre d'élèves influe sur l'attention de la classe ?



FIGURE 16 – Questions sur l'âge et le nombre des élèves

Ces deux questions, posées au début du questionnaire, affaiblissent malheureusement le lien entre le sondage et nos expériences. En effet, l'essentiel<sup>11</sup> des réponses du sondage concernait des classes de 15 à 35 élèves - tandis que nos classes en comptaient une quinzaine - et davantage des collégiens et lycéens que des jeunes adultes - comme c'était le cas dans nos expériences. Néanmoins, ce sondage nous a permis de découvrir des résultats intéressants sur la perception de l'attention par les professeurs, que nous résumons ici. Tout d'abord, il nous semble important de signaler que 100% de l'échantillon interrogé

9. lundi : 18.62% ; mardi : 30.58% ; mercredi : 17.53% ; jeudi : 24.14% ; vendredi : 6.71% ; samedi : 2.42%

10. Minimum : 0 an ; maximum : 41 ans ; moyenne : 15 ans et demi ; médiane : 15 ans ; écart-type : 9,7 ans.

11. Entre 25 et 35 : 60,1% ; Entre 15 et 25 : 31,5% ; Moins de 15 : 6,3%.

se dit « sensible aux variations d'attention de [ses] élèves lors des cours ». La sensibilité moyenne que les professeurs s'auto-attribuent sur une échelle discrète de 1 (« Pas du tout ») à 5 (« Très sensible ») est de 4,35.<sup>12</sup>

| <b>« Adoptez-vous des stratégies particulières pour remobiliser les élèves qui perdent le fil ? »</b>  |     |
|--|-----|
| Pas du tout  | 0%  |
| Oui, j'ai lu quelques ouvrages/articles sur le domaine et j'essaie d'appliquer les méthodes prescrites | 14% |
| Oui, sans doute quelques stratégies inconscientes  | 17% |
| Oui, j'essaie de remobiliser les élèves par le regard  | 63% |
| Oui, j'essaie de moduler le ton de ma voix   | 69% |
| Oui, j'essaie de basculer vers d'autres types de contenus (cours, exercices, animations...)            | 74% |
| Oui, j'essaie de me déplacer vers les foyers d'inattention   | 76% |

FIGURE 17 – Des stratégies particulières pour capter l'attention des élèves

Entre autres questions, nous avons demandé aux professeurs d'attribuer des scores de véracité de 1 à 5 ou « Je ne sais pas du tout » (1 = Complètement faux ; 5 = Très vrai) à 15 propositions sur l'attention. Voici leurs réponses, assorties de leurs écarts-types et pourcentages de « Je ne sais pas du tout ». Le gradient de couleur verte indique l'écart à la moyenne et on a mis en évidence les écarts-types supérieurs à 1 ainsi que les pourcentages de "Je ne sais pas du tout" supérieurs à 15%.

| PROPOSITION   | Score | $\sigma$ | %     |
|---|-------|----------|-------|
| L'attention d'un élève dépend de son état de <b>fatigue</b>   | 4.36  | 0.85     | 6.4%  |
| L'attention d'un élève dépend de son <b>intérêt</b> pour la matière enseignée                         | 4.01  | 0.91     | 2.0%  |
| L'attention de vos élèves dépend de l' <b>enthousiasme</b> que vous investissez dans le cours         | 3.91  | 0.95     | 4.6%  |
| L'attention d'un élève dépend de sa <b>faim</b>   | 3.88  | 0.93     | 4.8%  |
| L'attention d'un élève dépend de son <b>environnement</b> (voisins, fenêtre...)                       | 3.88  | 0.85     | 1.8%  |
| L'attention d'un élève dépend de la qualité de la <b>relation</b> que vous entretez                   | 3.85  | 0.98     | 3.6%  |
| L'attention de vos élèves augmente quand vous les <b>regardez</b>                                     | 3.74  | 0.93     | 3.7%  |
| L'attention de vos élèves augmente quand vous vous <b>déplacez</b>                                    | 3.50  | 1.00     | 6.8%  |
| Quand un <b>intervenant extérieur</b> s'exprime, les élèves sont plus attentifs                       | 3.29  | 1.05     | 3.5%  |
| <b>Plus</b> il y a d'enseignants, plus les élèves sont attentifs                                      | 2.89  | 1.20     | 18.8% |
| Juste après qu'on leur a rendu un <b>devoir</b> , les élèves sont plus attentifs                      | 1.89  | 0.96     | 6.9%  |
| S'il y a eu une activité la <b>veille</b> (kermesse, vide-grenier...), les élèves sont plus attentifs | 1.77  | 0.86     | 18.4% |
| S'il fait très chaud ou très froid, les élèves sont plus attentifs                                    | 1.74  | 0.97     | 7.9%  |
| La <b>capacité</b> d'attention est la même chez tous les élèves                                       | 1.45  | 0.96     | 0.3%  |
| L'attention d'un élève est la même quel que soit le <b>moment de la journée</b>                       | 1.37  | 0.88     | 0.0%  |

FIGURE 18 – Véracité de variables influançant l'attention des élèves

12. Ecart-type : = 0,67.

Enfin, nous avons interrogé les professeurs quant au « paramètre le plus crucial quand on étudie l'attention d'un élève » puis « d'une classe ».

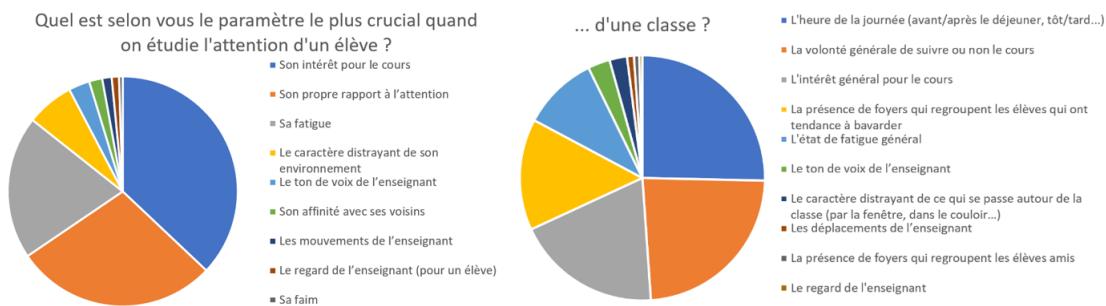


FIGURE 19 – Quel est le paramètre le plus crucial à propos de l'attention d'un élève ou d'une classe ?

Pour un élève comme pour une classe, si 93% (respectivement 95%) des professeurs se sont prêtés au jeu de choisir un critère parmi les 9 (respectivement 10) que nous avions proposés, beaucoup ont malgré tout préféré ajouter leur propre critère. Si la plupart du temps, c'est surtout pour signaler qu'il n'est pas possible de ne retenir qu'un critère, que tous entrent en jeu, que certains se recoupent, ou pour en sélectionner plusieurs d'un coup, parfois de nouveaux critères émergent.

Pour un **élève**, on peut citer :

- son environnement familial, s'il est paisible et valorise la formation apportée à l'école ;
- son hygiène de vie (sommeil, exposition aux écrans...), si elle favorise sa disponibilité à l'école ;
- sa tranquillité d'esprit, si elle lui permet de ne pas être mobilisé par autre chose que le cours ;
- sa capacité physiologique d'attention individuelle ;
- son niveau de compétence dans la matière ;
- la pédagogie de l'enseignant (continuité de l'enseignement au cours d'un cycle de formation, mise en œuvre d'activités variées, rapport de respect mutuel avec les élèves, volonté de faire des élèves des acteurs de leur formation et de les valoriser...)

Pour une **classe**, les mêmes critères reviennent, auxquels s'ajoutent :

- l'adaptabilité de l'enseignant et des activités à la diversité des profils des élèves ;
- la cohésion de la classe, si les élèves partagent une histoire scolaire ou extra-scolaire commune ;
- le nombre d'élèves dans la classe ;
- tantôt l'homogénéité, tantôt l'hétérogénéité du niveau scolaire du groupe ;
- l'identification des élèves bavards.

#### 4.4 Conclusion

L'entreprise de ce sondage a d'abord réussi à nous apporter un recul subjectif et expérimenté sur la façon d'étudier l'attention des élèves dans une classe.

La plupart des réponses possèdent des enjeux qui dépassent largement le cadre possible de nos expériences (ex : cadre familial ou physiologie de l'élève). Les méthodes d'analyses automatiques et statistiques ne resteront donc probablement qu'un seul type de cadre pour lire et tenter de prédire l'attention des élèves.

Ensuite, la pertinence de certains critères physiques a été confirmée : fatigue, faim et nombre d'élèves. Parmi ceux mesurables par notre méthode, nous retrouvons l'agitation ou la fatigue. D'autres, comme le sourire, ne sont pas présents.

Enfin, les professeurs confirment l'intuition d'évaluer l'intérêt des élèves pour le cours ou d'identifier les foyers de bavardage. Nous recommandons à un projet similaire d'approfondir en ce sens.

## 5 Discussion du projet : apports et limites

### 5.1 Apports du projet

Notre projet a d'abord cherché à **reproduire certains résultats de la bibliographie** : graphe de transition d'états et distribution des actions, en particulier ceux de Mirko Raca ([3]). Néanmoins, nous cherchions la complémentarité avec les études passées, notamment en entraînant une classe entière pendant environ 10 séances.

**L'apport principal** de notre projet a été de développer et tester un outil automatique de détection de l'attention. Il reste de nombreux points à améliorer et approfondir, mais nous avons tracé les contours d'une méthode efficace (du choix du logiciel jusqu'aux méthodes d'analyses pertinentes) en vue d'un outil de prédiction de l'attention avec les moyens d'un projet étudiant.

Le fait d'interroger les enseignants sur leur propre rapport à l'attention a apporté **un éclairage** intéressant à notre travail. Cependant, comme nous avons reçu les réponses assez tardivement, nous n'avons pas pu véritablement mettre leur expérience professionnelle au coeur de notre démarche : des développements ultérieurs de notre projet pourraient l'y intégrer davantage.

### 5.2 Limites des résultats, et prolongements possibles

La principale limite de ce PSC est bien sûr la **faible quantité de données récoltées**. Nous aurions pu gagner un facteur 4 (données exploitables sur dix séances au lieu de deux et demie) en étant plus directifs avec les participants (au risque de les braquer et de leur faire quitter l'expérience) et en anticipant quelques accidents. Mais pour réellement passer à l'échelle supérieure, il aurait fallu soit travailler avec de plus grands effectifs (ce qui implique de multiplier les caméras pour conserver une résolution suffisante, et de complexifier la mise en place), soit multiplier le nombre de classes.

Dans cette dernière éventualité, un groupe de PSC n'aurait pas le temps de mettre en place toutes ces expériences lui-même : il faudrait donc compter sur les enseignants pour assurer cette tâche eux-mêmes, ce qui est loin d'être évident (en 2e période, nous avons cherché à intervenir dans d'autres cours à l'X, mais n'avons pas obtenu l'accord des enseignants concernés). Il faudrait aussi disposer d'autant de caméras que de classes, alors que nous avions déjà du mal à emprunter une caméra toutes les semaines à l'association vidéo de l'X.

Au vu de cette insuffisance de données, et de la faible densité en élèves des cours que nous filmions, nous n'avons **pas pu étudier les phénomènes spatiaux** que nous voulions mettre en évidence (ex : influence des voisins). Dans notre analyse sur les seules données Wooclap (section 3.2.1), nous nous sommes donc tenus à des questions simples voire simplistes : évolution dans le temps de l'attention des élèves, influence de l'état de la classe sur celui d'un élève. Pour obtenir des résultats plus intéressants (ex : quantifier l'influence du plan de classe), il faudrait développer davantage les outils vidéo que nous avons commencé à créer et élargir le nombre de classes.

Nous avons rencontré de nombreuses **difficultés techniques** pour obtenir notre matériel semaine après semaine, trouver et faire tourner les bons logiciels de sondage et de traitement vidéo, manipuler des données lourdes, très fragmentées et mal étiquetées. Cela nous a fait perdre beaucoup de temps et a contrarié nos projets pour obtenir davantage de données.

Dans l'éventualité où notre projet serait repris par la suite, nous pourrions néanmoins fournir nos outils à nos successeurs et leur transmettre **les savoirs-faire que nous avons acquis dans l'expérimentation et le traitement de données**.

### 5.3 Discussion sur les outils et protocoles utilisés

**Les téléphones sont une trop grande source de distraction**, de l'avis conjoint des élèves et des enseignants : les boîtiers TurningPoint paraissent plus adaptés à tous égards. Une grande proportion d'élèves n'ont même jamais voulu participer à l'expérience dans ces conditions. D'autres rapportent d'ailleurs que le simple fait de s'interrompre pour répondre à un questionnaire est très perturbant : les boîtiers pallient en partie cet inconvénient, mais pour y remédier vraiment et obtenir davantage de coopération des élèves, on peut envisager d'espacer les mesures.

Nous aurions aimé **trouver une forme de contrat d'engagement** avec les sujets de l'expérience, ainsi qu'avec les partenaires extérieurs (ex : prêt de matériel de l'association vidéo). Une coopération sûre et régulière est fondamentale, et nos résultats ont pâti d'un engagement fluctuant de ces différents acteurs.

Du point de vue du protocole, la principale source d'erreur est clairement le "**clignotement**" des visages sur **OpenFace** évoqué figure 2. Sans outil informatique encore plus performant ou automatisé au maximum, la seule solution serait de *cropper* des visages individuels et de les traiter séparément : mais dans la mesure où un crop d'1h30 prend une

vingtaine d'heures à chaque ordinateur sur lequel nous avons travaillé, et où l'étiquetage et l'épuration des données *d'un seul cours* (marqueurs temporels, identification des visages sur OpenFace, conversion des fichiers csv en excel...) demande plus d'un week-end de travail, le traitement résultant serait une charge de travail assez lourde pour un groupe de PSC.

## 6 La vie de projet collectif

### 6.1 L'organisation en groupe

Au début du projet, nous avions participé aux **ateliers PSC** animés par Mme. Christelle Bisiau. Deux séances de 2h courant octobre nous ont permis de travailler :

- La qualité de la conduite du projet et du travail en équipe
- La préparation de nos échéances, notamment la proposition de PSC

Ces séances utiles demeurèrent toutefois, au début de notre projet où nous cherchions encore des prises d'accès, assez théoriques. L'animatrice nous a permis d'anticiper concrètement les étapes à venir. Restrospectivement, pouvoir expliquer oralement nos motivations et nos idées à une personne extérieure et sans formation scientifique spécifique était aussi extrêmement valorisable.

### 6.2 La coopération et la complémentarité

Nous avons pris énormément de plaisir à travailler ensemble et à nous compléter. Sous les conseils de Mme. Bisiau, nous nous sommes servis d'un diagramme de Gantt pour suivre l'évolution de notre projet. La répartition des tâches, elle, a toujours été naturelle et complémentaire. Chacun de nous nous retrouvions au moins par binôme sur les grandes figures du projet : Wooclap, OpenFace, boîtiers, traitement statistiques, modélisation mathématiques appliquées, traitement informatique, traitement Excel, expérience, sondage des élèves, sondage des professeurs.

#### PSC HSS 02



FIGURE 20 – Diagramme de Gantt du projet

## Table des figures

|    |  |    |
|----|--|----|
| 1  | Un programmeur du centre de recherche en éducation Hanwang présente le système "Class Care" à Beijing le 2 janvier 2019. Xue Yujie/Sixth Tone . . . . .  | 4  |
| 2  | OpenFace reconnaît le logo comme un visage ( <i>gauche</i> ). Le fichier excel correspondant fait alors la confusion entre les indices d'identification ( <i>en haut à droite : lorsque l'id 1 apparaît ; en bas à droite : quelques instants après</i> ). . . . . .   | 10 |
| 3  | Extraits de quelques comportements témoins . . . . .   | 12 |
| 4  | Relations entre les données extraites et les états prédicts : traitement par OpenFace (bleu), fonction d'encodage des données OpenFace en indice de comportement (jaune), nettoyage et organisation des données (orange), analyses statistiques pour la prédiction (vert). . . . .   | 14 |
| 5  | Exemple des variations d'incertitudes à $t = 00 : 20$ . . . . .  | 16 |
| 6  | Tableau résumant la réalisation des expériences. Vert : Expérience possible. Rouge : Expérience annulée. Noir : Expérience arrêtée . . . . .   | 17 |
| 7  | Qualité des réponses Wooclap et correspondance spatiale . . . . .  | 18 |
| 8  | Graphe de transition . . . . .   | 20 |
| 9  | Illustration du modèle Markov Caché . . . . .  | 21 |
| 10 | Histogramme d'évolution temporelle. Les deux séances du 21/12 ( <i>gauche</i> ) et 25/01 ( <i>droite</i> ). . . . .  | 22 |
| 11 | Influence de l'attention collective en HSS414b . . . . .   | 24 |
| 12 | Distribution des actions en fonction des états le 21/12 . . . . .  | 26 |
| 13 | Classification non-supervisée avec toutes les variables explicatives. On commence à distinguer deux communautés de données, bien que la distinction entre les deux blocs ne soit pas flagrante. La conclusion immédiate est que nous pouvons affiner notre recherche des variables pertinentes parmi ce lot initial de quinze variables. . . . . | 28 |
| 14 | Classification non-supervisée avec les 7 critères retenus (cf. 2.3.2). Les communautés se distinguent clairement ici. Ceci implique que la restriction à ces quelques variables se révèle pertinente. . . . .  | 28 |
| 15 | Résultats de l'analyse de variance à propos de l'inclinaison de la tête. A gauche la console R avec entre autres la p-value et les tests post hoc, à droite une représentation de l'inclinaison en fonction des états Wooclap, la moyenne est la barre noire . . . . .   | 31 |
| 16 | Questions sur l'âge et le nombre des élèves . . . . .  | 33 |
| 17 | Des stratégies particulières pour capter l'attention des élèves . . . . .  | 34 |
| 18 | Véracité de variables influançant l'attention des élèves . . . . .   | 34 |
| 19 | Quel est le paramètre le plus crucial à propos de l'attention d'un élève ou d'une classe? . . . . .  | 35 |
| 20 | Diagramme de Gantt du projet . . . . .   | 40 |

## Références

- [1] Toby Dragon al. Viewing student affect and learning through classroom observation and physical sensors. Intelligent Tutoring Systems, 2008.
- [2] Sébastien Poitrenaud Jean-Marc Bernard. L'analyse implicative bayésienne multi-variée d'un questionnaire binaire : quasi-implications et treillis de galois simplifié. Mathématiques et sciences humaines, tome 147.
- [3] Pierre Dillenbourg Mirko Raca, Lukasz Kidzinski. Translating head motion into attention - towards processing of student's body-language. 2015 in EDM.
- [4] Roland Tormey Mirko Raca and Pierre Dillenbourg. sleepers' lag - study on motion and attention. roceedings of the Fourth International Conference on Learning Analytics And Knowledge. P36-43. 2014.
- [5] Rana El Kaliouby Peter Robinson. Real-time inference of complex mental states from facial expressions and head gestures. 2004, Conference on Computer Vision and Pattern Recognition Workshop.
- [6] Yao Chong Lim Tadas Baltrušaitis, Amir Zadeh and Louis-Philippe Morency. Open-face 2.0 : Facial behavior analysis toolkit. IEEE International Conference on Automatic Face and Gesture Recognition, 2018.
- [7] Xue Yujie. Camera above the classroom. Sixth Tone, 26 mars 2019.

## A Modélisation mathématique en chaînes de Markov

### *Cadre du modèle*

La classe est modélisée par un réseau  $\Lambda$  comme un quadrillage de taille  $L$ . Donc  $\Lambda = \{1, 2, \dots, L\}^2$ . À chaque site  $i$  du réseau  $\Lambda$ , on associe un élève  $e_i$ . Ce dernier est égal à chaque instant à une valeur comprise dans  $\{0 = \text{"attentif"}, 1 = \text{"distrait par l'environnement"}, 2 = \text{"rêveur"}, 3 = \text{"esprit vide"}, 4 = \text{"distrait par l'expérience"}\}$ . On note  $S_\Lambda$  une configuration d'états d'attention.

Dans le formalisme des chaînes de Markov, une configuration  $S_\Lambda$  correspond à un état et l'espace d'états est  $E = \{0, 1, 2, 3, 4\}^\Lambda = F^\Lambda$ .

### *Probabilité des états mentaux : mesure de probabilité invariante*

Un élève  $e_i$  est représenté par une chaîne de Markov  $\{X_k\}_{k \in \mathbb{N}}$  irréductible, récurrente positive, d'espace d'états  $F$ , de matrice de transition  $P$  et de loi de probabilité invariante  $\mu$ .

Étant donné l'observation de  $(X_0, X_1, \dots, X_n)$ , qui sont les valeurs obtenues avec les réponses au sondage, on veut estimer  $\mu$  et  $P$  quand  $n \rightarrow \infty$ . Autrement dit, on cherche la probabilité que l'élève soit dans tel ou tel état d'attention ( $\mu$ ), et la probabilité que l'élève passe d'un tel état à un autre ( $P$ ).

Pour  $x \in F$ , on pose  $\mu^o(x; n) = \frac{1}{n+1} \operatorname{Card}\{0 \leq l \leq n : X_l = x\}$ .

D'après la loi forte des grands nombres, on sait que presque sûrement  $\mu^o(x; n) \rightarrow \mu(x)$  quand  $n \rightarrow \infty$ .