



HAL
open science

Tests statistiques et IRM cérébrales en classe de première S

Réjane Monod-Ansaldi, Grégoire Molinatti, Valérie Fontanieu, Daniel Devallois, Éric Sanchez

► To cite this version:

Réjane Monod-Ansaldi, Grégoire Molinatti, Valérie Fontanieu, Daniel Devallois, Éric Sanchez. Tests statistiques et IRM cérébrales en classe de première S. RDST - Recherches en didactique des sciences et des technologies , 2012, 5, pp.159-184. 10.4000/rdst.591 . hal-04286079

HAL Id: hal-04286079

<https://hal.univ-reunion.fr/hal-04286079v1>

Submitted on 15 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



RDST

Recherches en didactique des sciences et des technologies

5 | 2012
Varia

Tests statistiques et IRM cérébrales en classe de première S

Statistical tests and cerebral MRI in a scientific class in Première (Year 11)

Statistische Teste und Echoenzephalographie in der 11. Klasse mit Leistungskurs Mathe und Wissenschaft

Test estadísticos e IRM cerebrales en clase de 1^{re} S

Réjane Monod-Ansaldi, Grégoire Molinatti, Valérie Fontanieu, Daniel Devallois et Éric Sanchez



Édition électronique

URL : <https://journals.openedition.org/rdst/591>

DOI : 10.4000/rdst.591

ISSN : 2271-5649

Éditeur

ENS Éditions

Édition imprimée

Date de publication : 3 septembre 2012

Pagination : 159-184

ISBN : 978-2-84788-352-7

ISSN : 2110-6460

Ce document vous est offert par Bibliothèques de l'Université de La Réunion



Référence électronique

Réjane Monod-Ansaldi, Grégoire Molinatti, Valérie Fontanieu, Daniel Devallois et Éric Sanchez, « Tests statistiques et IRM cérébrales en classe de première S », *RDST* [En ligne], 5 | 2012, mis en ligne le 15 juillet 2014, consulté le 15 novembre 2023. URL : <http://journals.openedition.org/rdst/591> ; DOI : <https://doi.org/10.4000/rdst.591>



Le texte seul est utilisable sous licence CC BY-NC-ND 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.

Tests statistiques et IRM cérébrales en classe de première S

Réjane MONOD-ANSALDI

ENS de Lyon, IFÉ, EducTice - S2HEP

Grégoire MOLINATTI

Université de Montpellier 2-IUFM, LIRDEF

Valérie FONTANIEU

ENS de Lyon, IFÉ

Daniel DEVALLOIS

ENS de Lyon, IFÉ, EducTice - S2HEP

Éric SANCHEZ

ENS de Lyon, IFÉ, EducTice - S2HEP

RÉSUMÉ • Les images IRM fonctionnelles (IRMf) ont permis une avancée importante dans la compréhension du fonctionnement cérébral. Loin d'être de simples « photographies » du cerveau en fonctionnement, ces images résultent de protocoles de construction complexes et ont une signification statistique. La compréhension de leur nature statistique est difficile, mais nécessaire pour appréhender leur domaine de validité et construire une attitude critique face à leur utilisation, notamment dans les médias. Les représentations de ces images construites par les élèves ne prennent généralement pas en compte cette dimension statistique. Nous présentons et analysons une séquence d'apprentissage innovante mise en œuvre pour favoriser la compréhension de ce type d'image. Il s'agit d'une séance de modélisation comprenant le paramétrage d'IRMf à l'aide du logiciel EduAnatomist, précédée d'une activité d'appropriation de la notion de test statistique (test T). Nos résultats font apparaître les difficultés inhérentes à la prise en compte par les élèves de l'aspect statistique des images biologiques. Nous proposons des pistes didactiques pour ouvrir, au moins en partie, certaines des boîtes noires qui correspondent à la construction de ce type d'image.

MOTS-CLÉS • Neurosciences, imagerie par résonnance magnétique, modélisation, statistique, représentation, résolution de problème.

À grand renfort d'images cérébrales, un sujet proposé dans le très suivi journal télévisé de 20 heures nous décrit comment les neurosciences expliquent la genèse du sentiment amoureux. Mais que représentent au juste ces images largement médiatisées ? Comment sont-elles construites ? Quelles sont leurs limites en termes de validité ?

« Quand on tombe amoureux, douze aires du cerveau s'activent et travaillent ensemble pour libérer des substances euphorisantes [...] un phénomène que l'on observe grâce à l'imagerie médicale, le coup de foudre se définit donc scientifiquement ».

(France 2, journal télévisé, 20 h, 9 novembre 2010)



Si les images fonctionnelles permettent l'étude d'un certain nombre de fonctions cérébrales et de comportements (sensibilité, motricité, mémoire, apprentissage, spiritualité, sexualité, agressivité, etc.), l'histoire sociale des sciences biologiques nous invite cependant à une certaine vigilance en particulier vis à vis des réductionnismes souvent associés à la biologisation des questions psychologiques et sociales. Gould (1983) a ainsi montré comment l'anthropologie biologique avait pu alimenter des idéologies racistes ou sexistes. Loin de rester confinées à l'histoire des sciences, les implications sociales des neurosciences modernes soulèvent des interrogations éthiques bien contemporaines, telles que les tendances sexistes de certaines recherches mobilisant des données de neuro-imagerie (Clément, 1997 ; Vidal & Benoît-Browaëys, 2005) ou les questions qu'elles posent en termes de libre arbitre et de déterminisme biologique (Malabou, 2004). Les citoyens européens, qui ont participé à la consultation publique *Meeting of Minds* en 2006 ont ainsi soulevé des questions concernant la réglementation et le contrôle des recherches sur le cerveau, l'information sur ces recherches (fiabilité, transparence, impartialité, etc.), l'éducation et la sensibilisation des non-spécialistes. Ils se sont questionnés sur la définition de la normalité et la prise en compte de la diversité des personnes (notamment en termes de maladies mentales), sur la pression des intérêts économiques (neuro-marketing), et sur l'égalité d'accès aux soins. Autrement dit, une réflexion sur l'enseignement des neurosciences doit prendre en compte des questions de culture scientifique, relatives à la capacité des citoyens à se saisir des enjeux de société associés à ce domaine de connaissances scientifiques.

C'est pourquoi l'enseignement des neurosciences doit certes viser des objectifs de connaissances, mais aussi prendre en compte la dimension épistémologique. De ce point de vue, la maîtrise des processus d'iconogenèse est essentielle à l'iconolecture, c'est-à-dire à la compréhension des images (Clément, 2001), même si elle ne paraît

pas suffisante pour en appréhender les implications sociales en lien avec des idéologies et des valeurs. Le travail que nous présentons ici¹ se préoccupe de cette première étape de compréhension de la construction des images cérébrales fonctionnelles, qui mobilise des protocoles complexes, des modèles statistiques et des procédés de représentation d'une structure en volume (orientation des plans de coupe, taille, résolution, contraste) et conditionne leur domaine de validité.

1. Les IRM fonctionnelles, des images difficiles à comprendre

1.1. Les IRM fonctionnelles cérébrales, des modèles statistiques du fonctionnement cérébral

L'IRM consiste à observer les tissus biologiques à travers les propriétés magnétiques de l'un de leurs constituants majoritaires, le noyau d'hydrogène. Lors d'une acquisition, chaque unité de volume du cerveau ou voxel est caractérisée par l'intensité du signal magnétique mesuré, ce qui permet de reconstruire des images dans les trois plans de l'espace. L'IRM anatomique permet de distinguer la nature des tissus par la détection de l'eau peu mobile c'est-à-dire intracellulaire (signal élevé pour la substance grise et faible pour les os par exemple) et donc d'obtenir des images numériques, d'une précision inférieure au millimètre, où l'on peut repérer les principales structures cérébrales (cortex, corps calleux, liquide céphalo-rachidien, noyaux gris centraux, etc.). Des traitements mathématiques permettent d'améliorer les images en corrigeant, par exemple, les éventuels mouvements, ou encore en les « normalisant », c'est-à-dire en les ramenant à un cadre spatial de référence dit de Talairach pour pouvoir comparer différents sujets.

L'IRM fonctionnelle utilise les propriétés paramagnétiques de la désoxy-hémoglobine et permet l'observation en temps réel des variations de l'oxygénation du sang et des débits sanguins cérébraux locaux, en lien avec le fonctionnement cérébral. L'activité cérébrale s'accompagne en effet d'un enrichissement en oxygène des régions mises en jeu qui se traduit par une augmentation locale du signal IRM, appelé BOLD (Blood Oxygen Level Dependent). La construction même des images cérébrales fonctionnelles est étroitement dépendante du paradigme expérimental utilisé pour activer le cerveau en isolant l'opération cognitive que l'on veut étudier du reste de l'activité cérébrale. Plusieurs protocoles ou paradigmes de construction d'images cérébrales sont utilisés. Le plus simple est sans doute le paradigme en bloc dit de soustraction ou de différence (figure 1). On réalise une série d'acquisitions du signal magnétique au cours d'une condition de référence (par exemple sans stimulus) et une autre série d'acquisitions en condition dite cible (par exemple avec stimulus). Les IRM brutes obtenues pour chaque mesure font correspondre à chaque voxel une couleur symbolisant l'importance du flux sanguin. Les IRM moyennées représentent, pour chaque voxel, les moyennes entre plusieurs mesures du flux

1 Ce travail a été réalisé au sein de l'équipe ACCES de l'ENS de Lyon-IFÉ.

sanguin réalisées dans les mêmes conditions. Pour obtenir une IRM fonctionnelle, un test de différence statistique des signaux d'acquisition entre les deux conditions (test T de Student par exemple) est appliqué voxel à voxel. Dans l'IRMf construite, chaque voxel se voit donc affecté d'une valeur mathématique (la statistique de test) résultant du test de la différence entre les deux séries correspondant aux deux conditions d'acquisition (comparaison des moyennes des deux séries avec le test de Student). À ces valeurs correspond une probabilité de risque de conclure à tort à une différence significative. Ces probabilités sont ici traduites selon une palette de couleurs différenciées (spectres de couleurs chaudes – froides) permettant de visualiser le niveau de signification statistique de la différence pour chaque voxel. Autrement dit une image cérébrale fonctionnelle est une construction mathématique spatialisée à laquelle il serait plus judicieux de donner le nom de carte fonctionnelle statistique. L'image fonctionnelle est généralement superposée à une image anatomique pour pouvoir localiser les régions cérébrales spécifiquement activées au cours de la fonction cognitive étudiée.

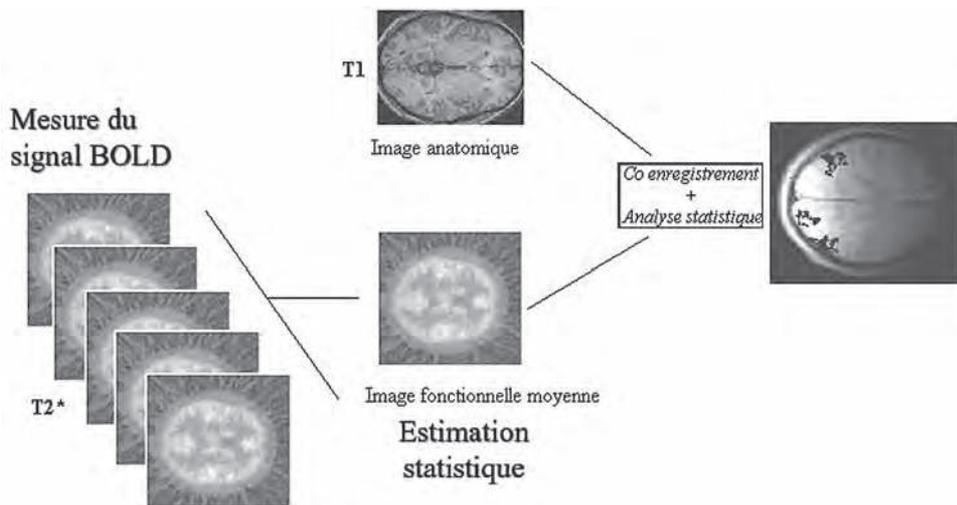


Fig. 1 : protocole de construction des IRM cérébrales fonctionnelles.
D'après Segebarth C. & Ziegler A. (2004).

La réalisation d'IRMf correspond donc à une modélisation du fonctionnement cérébral, au sens de Bunge (1975) qui définit les modèles comme des outils conceptuels élaborés par le chercheur, permettant de faire le lien entre théorie et réel. Sa construction reprend d'ailleurs les étapes de la production des images scientifiques décrites par Jacobi (1987) comme une « focalisation progressive qui permet de transformer un réel complexe en un observable simple sur lequel le chercheur pointe toute son instrumentation pour enregistrer un changement, ou mesurer l'amplitude (ou la fréquence) d'un signal » et qui « détermine très largement le type d'image produit ». Des allers retours entre le modèle et le réel qu'il représente sont donc nécessaires pour permettre au chercheur d'aborder globalement l'image produite comme une visualisation simplifiée du réel dans un

but heuristique (Latour, 1985). Dans le cadre d'un usage scolaire d'IRMf, ces allers retours ne sont pas possibles pour les élèves si le processus de construction des images est éludé. Il y a là, selon nous, un obstacle majeur à l'interprétation correcte de ces images par les élèves et donc un enjeu didactique fort.

1.2. Conceptions d'élèves sur les neurosciences et les IRMf

De nombreux travaux de didactique des sciences, dus notamment à Clément et à ses collaborateurs, ont documenté les conceptions de différents publics sur le système nerveux et son fonctionnement (Mein & Clément, 1988 ; Clément, 1994, 1997 ; Forissier & Clément, 2000 ; Savy, 2005, etc.). Ainsi, les conceptions des non-spécialistes sont marquées par une persistance des modèles dualistes concernant les rapports entre le cerveau et l'âme, entre la matière et la pensée. Ces conceptions sont aussi dominées par la valorisation du déterminisme génétique du développement et du fonctionnement cérébral, et particulièrement de l'intelligence (Van Blyenburgh, 1989). Le concept de plasticité, susceptible de rendre compte des phénomènes d'épigenèse cérébrale, est largement ignoré alors que c'est un concept structurant des neurosciences modernes. Le cerveau revêt pour les non-spécialistes un caractère mystérieux presque fascinant qui repose en partie sur son extrême complexité. Car bien que constitué de « parties » qui sont le siège de fonctions distinctes, les personnes interrogées relèvent souvent « le mystère des parties inexplorées qui renvoient à un imaginaire topographique propre à favoriser les évocations d'explorations et de découvertes » (Le Marec, Boucher & Hiard, 1992).

Chez les enfants, le cerveau joue un rôle important dans la « commande » ou le « contrôle » plus que dans les émotions par exemple. Des analogies avec les fils, l'existence d'une mémoire, d'une « unité centrale » sont évoquées. Ces conceptions reposent sur un modèle dominant de cerveau/ordinateur (Mein & Clément, 1988 ; Hergueta, Lemire & Girault, 1997). Un autre modèle est parfois convoqué selon ces auteurs, modèle qu'ils qualifient de « processus entéroïde » puisqu'il mobilise des analogies avec les intestins. Nous avons mené en 2001 puis en 2005 des entretiens semi directifs visant à préciser les conceptions (le déjà-là conceptuel) des adolescents sur le cerveau. Ces entretiens ont montré que le modèle du cerveau muscle qui commande, qui actionne, est largement mobilisé. Le cerveau est d'autant plus développé que l'on s'en sert. Cette propriété, souvent associée aux mécanismes d'apprentissage, traduit la conception d'un cerveau adaptable et modulable. D'ailleurs, si elle n'était pas évoquée dans les entretiens de 2001, la neuroplasticité est souvent citée par les lycéens en 2005². Une autre constante de la cartographie

2 La plasticité cérébrale est souvent associée aux mécanismes d'apprentissage (« Plus on apprend, plus ça se connecte » ; « Notre cerveau évolue au cours de notre vie, il y a des nouvelles connexions, les neurones se renouvellent » ; « Pour apprendre, il faut faire des connexions, des synapses »). Sans que soit nécessairement proposé un processus d'épigenèse, plusieurs lycéens avancent qu'il faut « entretenir » le cerveau, l'« entraîner », le « faire travailler » pour qu'il se développe « c'est comme un muscle ». Enfin l'utilisation partielle des capacités du cerveau reste une constante des conceptions et on peut confirmer que certains adolescents y conçoivent un support de l'épigenèse (« À la base,

des conceptions des adolescents sur le cerveau est celle des fameux « x % du cerveau utilisés ». Bien que sans aucun fondement scientifique, il semble que pour une majorité d'adolescents, le cerveau fonctionne de manière modulaire avec certaines zones au repos alors que d'autres sont en activité. Il est possible que ce modèle soit en relation avec les représentations de neuro-imagerie (Molinatti, 2007).

Pour autant, peu de données empiriques sont disponibles sur les conceptions des élèves relatives aux données de neuro-imagerie. Clément & Fisseux (1999) dans leurs travaux sur l'imagerie médicale, ont montré que les élèves confondent radiographies et photographies. Le travail de Mafféo (2007) concernant le statut et le rôle des images cérébrales dans l'enseignement secondaire fait état de l'écart entre la signification de ces images dans le monde « médico-scientifique » et leur signification dans le monde de l'enseignement. Les images scientifiques complexes telles que les IRM fonctionnelles ne peuvent être signifiantes sans les informations concernant leurs protocoles de constructions et sans les éléments de légende permettant de les interpréter. Mobiliser ces images pour l'enseignement sans travailler sur leurs conditions de production, et donc d'interprétation, c'est restreindre leur compréhension et favoriser leur instrumentation ou le réductionnisme de leur interprétation (Mafféo, 2007). Pour notre part, nous avons exploré les conceptions de 30 élèves de 1^{re} S suite à une séance de TP, comprenant la visualisation d'IRMf avec le logiciel EduAnatomist pour la recherche des régions cérébrales spécifiquement activées par la stimulation tactile de l'index droit. Alors que l'enseignant avait présenté le principe statistique de construction des IMRf, un questionnaire proposé dans la semaine suivant la séance a révélé que, pour 60 % des élèves, une IRMf est « une représentation de l'activité du cerveau à un moment donné ». C'est aussi une « radiographie du cerveau », un « enregistrement de l'état du cerveau à un moment donné » pour la moitié des élèves ou même une « photographie du cerveau » pour 17 % d'entre eux. Ces quatre propositions fortement retenues, ne tiennent pas compte du protocole soustractif de construction des images et du traitement statistique appliqué aux données issues des mesures. Seuls 37 % des élèves retiennent « une construction réalisée par des chercheurs à partir de plusieurs mesures », 30 % « une représentation des différences statistiques d'irrigation du cerveau entre deux séries de mesures », et 10 % « une comparaison entre deux états différents ». Ainsi la simple présentation du protocole de construction des images ne suffit-elle pas à modifier les représentations des élèves, pour qui les IRM restent avant tout des « radiographies » ou des « photographies » du cerveau.

rien n'est établi. D'ailleurs Einstein disait que le cerveau était utilisé à 70 %). Pour interpréter cette évolution des conceptions d'adolescents concernant le cerveau il faut préciser que les programmes officiels d'enseignement sont marqués par l'intégration de la neuroplasticité et de l'épigénèse dans le développement et le fonctionnement cérébral, entendus comme facteurs d'individuation.

1.3. Difficultés potentielles à la compréhension des IRMf par les élèves

Une des difficultés de compréhension et d'interprétation des IRMf par les élèves relève de la représentation du cerveau, objet à trois dimensions, par trois images en deux dimensions : des coupes cérébrales, réalisées selon trois plans perpendiculaires entre eux (voir, par exemple, figure 2). Sur ces trois coupes, les voxels deviennent des pixels colorés, et un même voxel peut être représenté simultanément dans les trois plans, donc dans les trois images observées. L'articulation entre les images 2D et les structures 3D représente effectivement un obstacle, comme l'a souligné Clément (1996).

La compréhension du test statistique mobilisé dans la construction des IRMf présente des difficultés. Pour Y. Girault et M. Girault (2004), « l'obstacle le plus fondamental [à l'utilisation de modèles statistiques] réside dans l'attitude intellectuelle : il faut oublier les cadres de pensées du déterminisme pour s'ouvrir à d'autres. Or, compte tenu des réussites spectaculaires du déterminisme, et peut-être de sa simplicité, il s'est tellement imposé qu'il n'a plus été considéré comme l'hypothèse, mais véritablement comme un dogme. » (p. XI-XII). Ces auteurs insistent aussi sur la difficulté des techniques de la statistique, qui font appel à des notions originales nécessitant une étape de familiarisation. Le fait que face aux conceptions déterministes, le « hasard » soit invoqué, terme chargé de sens divers, parfois mythiques ou ésotériques, et par la suite très imprécis, est un obstacle en lui-même. Même si le terme plus neutre d'aléatoire est utilisé, le changement de vocabulaire ne suffit pas pour supprimer la difficulté de compréhension de la signification profonde des calculs qui seront effectués.

L'articulation de la notion de test statistique avec les questions biologiques, n'est pas aisée non plus. En biologie, l'utilisation de tests statistiques est souvent liée à l'expérimentation sur des phénomènes variables. En effet, mesurer une grandeur variable, c'est établir un échantillon de données représentant toutes les valeurs que cette grandeur pourrait prendre. Plus le nombre de mesures est important, plus l'idée que l'on peut se faire de la grandeur mesurée est précise. Mais lorsque l'on veut comparer cette grandeur dans deux conditions différentes, ce qui est un objectif fréquent de la démarche expérimentale en biologie, il faut discerner la variabilité de la grandeur dans chacune des conditions, de la variabilité entre les deux conditions, et déterminer si la différence entre les deux conditions est significative. C'est l'objet du test statistique de Student, qui évalue la probabilité de rejeter à tort l'hypothèse nulle, c'est-à-dire l'hypothèse de l'absence d'effet, de l'absence de différence entre les deux conditions. C'est ainsi l'hypothèse nulle qui est testée. On rencontre ici une difficulté supplémentaire car généralement, l'hypothèse posée en biologie est inverse : si l'on fait l'expérimentation, c'est bien pour explorer ou mettre en évidence un effet d'un facteur isolé sur la variable étudiée.

Le travail que nous présentons dans ce compte-rendu d'innovation est une première exploration de situations d'apprentissage permettant aux élèves de

construire une signification statistique des IRM fonctionnelles. Il se focalise sur les procédés d'obtention de ces images, en particulier sur le rôle central que joue le test de Student dans leur construction, dans le but d'outiller leur lecture par les élèves. Il prend également en compte le nécessaire travail sur la représentation du volume cérébral par le choix d'un logiciel qui permet à la fois d'explorer le volume cérébral, de repérer des liens entre les trois plans de coupes et de paramétrer des images fonctionnelles. Nous présentons les deux séances expérimentées avec les élèves et leur analyse.

2. Séquence d'enseignement sur la modélisation statistique dans les IRMf

Deux séances de travaux pratiques de 1^{re} S de deux heures ont été conçues dans le but de travailler sur le caractère modélisant des IRM fonctionnelles. La première a été consacrée au problème de la comparaison de deux séries de mesures d'une grandeur variable et avait pour objectif d'aborder la notion de test T ou test de Student. Dans le cadre de l'étude du fonctionnement du système nerveux, le problème était de déterminer si écouter de la musique pouvait influencer le temps de réaction motrice à un stimulus visuel. La deuxième séance a consisté à travailler sur des IRM fonctionnelles à l'aide du logiciel EduAnatomist, pour déterminer quelles régions cérébrales sont activées lors de la stimulation tactile de certaines zones du corps.

2.1. Séance 1 : comparer deux séries de mesures d'une grandeur variable

Tâches proposées aux élèves

Après avoir proposé une hypothèse concernant l'effet de la musique sur leur temps de réaction, les élèves ont utilisé le logiciel *Réaction*³, qui permet de mesurer des temps de réaction lors d'apparitions d'un stimulus visuel. Travaillant par binôme, ils ont effectué deux séries de 30 mesures du temps de réaction d'un des deux élèves, sans fond sonore puis en écoutant de la musique. Les données obtenues (comprises entre 125 et 687 ms) ont été saisies dans une feuille de calcul *Excel*, les moyennes ont été calculées pour les deux conditions. Les élèves ont fourni une première réponse concernant l'effet de la musique sur leur temps de réaction et ont partagé leurs résultats au tableau. D'un élève à l'autre, les résultats étaient très différents, mais le problème de l'effet de la musique était posé à l'échelon individuel, pour chaque élève ayant réalisé les mesures. La différence entre les moyennes dans les deux conditions variait entre 0 et 63 ms. Un débat s'est engagé

3 *Réaction*, logiciel INRP créé par F. Tilquin, 2007, En ligne : <<http://accs.inrp.fr/accs/ressources/neurosciences/temps-de-reaction-investigation-variabilite-et-traitements-statistiques-des-donnees/ressources-1/le-logiciel-reaction-1/>>.

sur la façon de trancher quand les moyennes n'étaient que « peu » différentes. Le problème émanant de la situation était donc de déterminer si l'écart entre les deux séries de mesures était significatif et permettait de conclure. L'enseignant a alors présenté aux élèves le principe du test de Student et leur a fourni une feuille *Excel* permettant d'appliquer ce test à leurs propres mesures et d'afficher la probabilité associée à la statistique de test obtenue (p-value) ainsi que la valeur de la statistique de test au-delà de laquelle on conclut à une différence significative pour un niveau de risque donné (niveau qu'ils pouvaient faire varier). Il nous paraissait important de proposer aux élèves un outil de calcul qui ne pose pas de difficulté opératoire, et qui leur permette de manipuler le modèle statistique proposé. Munis de cet outil, les élèves ont exploité leurs propres résultats pour répondre aux questions suivantes : « *Indiquez si l'environnement modifie le temps de réaction. Argumentez votre réponse.* » et « *La différence entre les deux séries de mesures est-elle significative au seuil de confiance 95 % ? Argumentez votre réponse.* ».

Les conditions de validité du test n'ont pas été vérifiées avec les élèves. Les données de précédentes séries de mesures avaient été étudiées avant la préparation de la séance, témoignant de faibles distorsions vis-à-vis des conditions à réunir.

Utilisation du test statistique par les élèves

L'implication générale des élèves a été bonne durant la première séance. Les mesures du temps de réaction ont pris un caractère ludique et le fait de travailler sur leurs propres résultats pour résoudre un problème qui les concerne directement a sans doute favorisé la dévolution du problème. Les données obtenues ont donné lieu à une analyse statistique (Fontanieu *et al.*, 2010).

Sur les quinze binômes ayant participé à la première séance, une très grande majorité (13) a utilisé correctement la feuille *Excel* prédéfinie pour calculer la statistique de test correspondant à ses mesures. Les deux binômes qui n'ont pas réussi, n'ont pas répondu non plus aux questions : l'un a rencontré un problème technique avec l'ordinateur, l'autre était un trinôme, dont un élève avait dû s'absenter, ce qui a complètement démobilisé le groupe. L'observation durant la séance a montré que certains groupes utilisaient à plusieurs reprises la feuille « loi de Student » pour évaluer le risque de se tromper associé à une statistique de test, ou l'inverse.

Sur les treize binômes ayant répondu à la question « *2. Indiquez si l'environnement modifie le temps de réaction. Argumentez votre réponse.* », deux donnent des réponses contradictoires avec le résultat du test, sans le citer et en l'ayant cependant calculé, et onze binômes produisent une réponse cohérente avec les résultats du test. Cette cohérence n'indique cependant pas que le résultat du test a été utilisé pour élaborer la réponse. En effet, trois groupes n'argumentent pas leur conclusion, trois binômes comparent les moyennes, deux le « temps de réaction avec et sans musique », et un « le témoin et l'expérience ». On peut remarquer que ces six derniers binômes agissent ici comme de coutume en sciences de la vie, où il nous semble que la

significativité statistique des différences observées entre les mesures réalisées en classe n'est presque jamais questionnée.

Seuls quatre binômes font référence au test statistique dans leur réponse. Leurs textes, présentés dans le tableau 1, suggèrent qu'ils ont compris l'utilité de ce test, même si la formulation retenue n'est pas celle habituellement utilisée en statistique. Ainsi, le raisonnement statistique proposé n'est pas facile à assimiler et rares sont les binômes qui l'ont acquis au cours de cette unique séance. La maîtrise des concepts liés à l'utilisation d'un test statistique demanderait sans aucun doute un travail plus approfondi.

Binôme	Texte écrit par les élèves
(2A)	« D'après les résultats que nous avons obtenus, il nous est impossible d'affirmer que l'environnement modifie le temps de réaction. En effet, nous avons obtenu une légère différence dans les tests effectués avec et sans musique. Cette différence n'est pas significative, le pourcentage d'erreur est alors trop élevé. »
(3A)	« Selon notre travail de mesure, l'environnement modifierait le temps de réaction. En effet la différence de moyenne entre nos deux séries de mesures est de 14 ms moins importante pour la série effectuée sans musique. De plus notre valeur de t calculée est de 3,14 et correspond à un risque d'erreur de 0,27 %. On peut donc dire que d'après nos mesures, on peut affirmer avec quasi certitude que l'environnement modifie le temps de réaction. »
(1B)	« On ne peut pas réellement déterminer si l'environnement modifie le temps de réaction car le risque de se tromper sur la différence est de 72 %, ce qui est élevé. »
(3B)	« Le risque est de 7,24 %. Cette valeur est supérieur [sic] au 5 % admis par les scientifiques, on peut donc admettre que l'environnement ne modifie pas le temps de réaction. »

Tableau 1 : réponses des binômes ayant mobilisé le test dans leur argumentation à la question « 2- Indiquez si l'environnement modifie le temps de réaction. Argumentez votre réponse. »

À la question plus technique « 3. La différence entre les deux séries de mesures est-elle significative au seuil de confiance 95 % ? Argumentez votre réponse. », neuf binômes sur quinze proposent des réponses cohérentes avec les résultats du test. Cependant seulement sept binômes témoignent d'une compréhension de la question et argumentent leur réponse en utilisant le principe du test. On observe fréquemment une confusion entre la statistique de test, calculée à partir des mesures, et la valeur correspondant au niveau de risque fixé. La formulation de la question dans le sens de la confiance au lieu de celui du risque, qui a été utilisé durant la présentation du test par l'enseignant a sans doute encore complexifié la compréhension de cette question par les élèves. Elle pourrait être remplacée par « *Pouvez-vous affirmer avec un risque de vous tromper inférieur à 5 %, que la différence entre vos deux séries de mesure est significative ?* ».

2.1. Séance 2 : modifier les paramètres du modèle statistique pour appréhender la signification des IRM

Cette séance mobilise la banque de données de neuro-imagerie *Neuropeda* et le logiciel de visualisation de ces images *EduAnatomist*, développés par l'équipe ACCES (ex INRP, ENS de Lyon-Ifé)⁴.

Tâches proposées aux élèves

Après avoir analysé des données cliniques (conséquences fonctionnelles d'accidents vasculaires cérébraux) permettant de faire l'hypothèse d'une régionalisation de certaines fonctions cérébrales, les élèves ont utilisé *EduAnatomist*, pour visualiser et paramétrer deux images IRM fonctionnelles de la banque *NeuroPeda*, permettant d'étudier les régions cérébrales activées par la stimulation tactile de différentes zones du corps (figure 2). La tâche consistait à sélectionner les bornes d'affichage déterminant les voxels affichés sur une échelle de significativité, allant sur le logiciel de 0 à 100. La borne inférieure définit la valeur à partir de

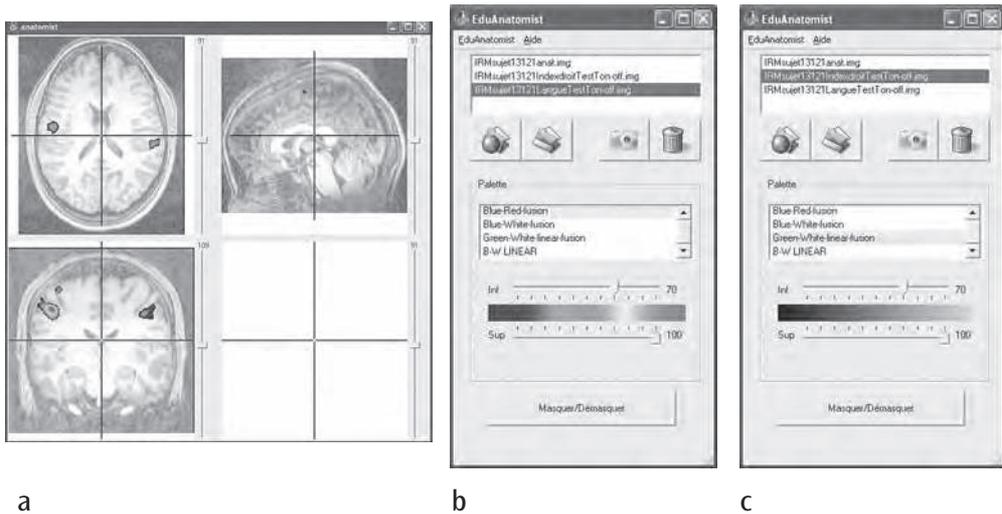


Fig. 2 : interface du logiciel *EduAnatomist*, affichant une IRM anatomique et deux IRMf du même individu (zone activée par la stimulation tactile de l'index droit, zone activée par la stimulation tactile de la langue).

Copies d'écran du groupe 1B, fenêtre d'affichage des images (a) et fenêtres de paramétrage des images concernant la stimulation de la langue (b) et de l'index (c).

4 Ce projet a reçu le soutien du ministère de l'Éducation nationale (projet Schene, 2007). Les enseignants de sciences de la vie et de la Terre qui ont participé à ce développement sont Sandrine Beaudin, Julien Cartier, Éric Chauvet, Anne Florimond, Marie-Claire Garnier, Françoise Jauzein, Sandrine Lecoq, Christine Montixi, Nathalie Noris, Aude Richter, Michèle Ternaux et Yannick Vilain. Le développement a été coordonné par Grégoire Molinatti de 2003 à 2009, puis par Guillaume Rami, responsable de l'équipe de Marseille, et Philippe Daubias. En ligne : <http://acces.inrp.fr/acces/ressources/neurosciences/Banquedonnees_logicielneuroimagerie>.

laquelle les voxels sont colorés. Plus la valeur de la borne inférieure est élevée, plus l'exigence de significativité est importante et plus la spécificité de la zone colorée est importante. La borne supérieure définit la valeur à partir de laquelle les voxels sont affichés avec la couleur extrême supérieure de la palette, si bien que ce sont les valeurs comprises entre les deux bornes qui sont graduellement colorées selon l'échelle de la palette. Lorsque la borne supérieure est à 100, toutes les valeurs supérieures à la borne inférieure sont discriminées par la couleur. Si elle est trop peu élevée, des voxels de faible significativité sont affichés comme les voxels de forte significativité. La palette de couleur est automatiquement redistribuée entre les valeurs des bornes inférieure et supérieure.

Lors de la présentation du logiciel, l'enseignant définissait l'échelle présente sur l'interface (figure 2, b et c) comme représentant la statistique de test, associée au test de significativité de la différence entre les deux séries de mesures correspondant à la stimulation et à l'absence de stimulation tactile de chaque zone du corps étudiée. Il précisait que plus la couleur était chaude ou intense, plus la statistique de test était elle-même élevée et plus la différence était significative.

Chaque binôme devait identifier deux zones (par exemple, celles correspondant à la stimulation tactile de l'index et du pied), et produire au moins une diapositive présentant le résultat de ce travail. La diapositive était ensuite envoyée par MSN, un outil de clavardage, à un autre binôme, ayant travaillé sur la stimulation de l'index et d'une autre zone du corps (par exemple, la langue) et une collaboration à distance s'engageait pour produire une diapositive commune destinée au site Web de l'établissement et permettant de visualiser les trois régions cérébrales étudiées. Durant cette collaboration par clavardage et échange de fichiers, les élèves n'étaient pas en mesure de regarder directement ce qui était affiché sur leurs écrans respectifs et devaient formuler et parfois justifier leur choix.

Paramétrage des images par les élèves

Les binômes du premier demi-groupe (1A à 8A) avaient à leur disposition un nombre important de fichiers images et devaient choisir entre des IRM brutes, des images moyennées et des IRM statistiques correspondant à un test T. Ce travail s'est révélé très difficile et chronophage et les élèves n'ont pas eu le temps d'échanger sur leurs images. Autrement dit, bien que le nom donné aux images spécifie si ce sont des images moyennes ou des IRM statistiques, il semble se confirmer que la « mise en image et en espace » de traitement mathématique n'est pas saisie spontanément par la majorité des élèves. Les données récoltées pour ce groupe sont donc moins riches et il est difficile d'y retrouver la mobilisation des éléments de statistique travaillés pendant la première séance.

Nous avons pris la décision de restreindre le choix des images pour le deuxième demi-groupe (1B à 8B), de manière à libérer du temps pour leur paramétrage et pour le travail collaboratif entre binômes. Les bornes statistiques retenues et

le nombre de zones cérébrales spécifiques effectivement visibles sur les images produites par les élèves sont présentés dans le tableau 2.

Binôme	1 ^{re} séance : apprentissage du test T		2 ^e séance : paramétrage des IRMf	
	Utilisation du test pour répondre au problème (question 2)	Maîtrise opérationnelle du test pour répondre à la question technique (3)	Paramétrage des bornes statistiques des 2 images fonctionnelles	Nombre de zones cérébrales visibles sur l'IRM
1B	test compris	+	70-100 70-100	2
2B	test non utilisé	+	32-100 32-100	0
3B	test compris	+	49-100 75-100	2
4B	absent	absent	70-100 70-100	2
5B	test non utilisé	+	50-100 78-100	2
6B	test non utilisé	-	44-95 44-95	2
7B	test non utilisé	-	56-100 69-100	2
8B	test non utilisé	-	30-100 30-100	0

Tableau 2 : appropriation de la notion de test T lors de la première séance et paramétrage statistique des IRM lors de la deuxième séance (deuxième groupe, 8 binômes)

Six binômes sur huit ont réussi à afficher les deux zones cérébrales recherchées, en sélectionnant des bornes statistiques correspondant à une significativité plutôt élevée. Parmi eux se trouvent les deux binômes qui avaient montré une compréhension suffisante du test T lors de la première séance (1B et 3B). Dans sa réponse à la question finale, « *Utilisez votre travail pour revenir sur les hypothèses proposées* », le binôme 3B reprend d'ailleurs la notion de risque en écrivant : « *On a fait le choix de sélectionner un risque peu élevé car c'est entre les valeurs 50 (ou 75 inf) à 100 sup que les zones les plus irriguées sont les mieux localisées* ». Les choix effectués par ces élèves sont donc justifiés, explicites et clairs. La remobilisation du vocabulaire concernant le risque de se tromper utilisé lors de la première séance, indique que le lien entre les deux activités a bien été saisi par les élèves et qu'ils ont su transférer dans cette séance des connaissances statistiques acquises lors de la séance précédente. Les zones recherchées sont caractérisées par leur irrigation (*les zones les plus irriguées*), et non en termes de signification statistique de la différence entre deux conditions, ou de spécificité par rapport au stimulus appliqué, ce qui

confirme l'importance accordée par ces élèves au corrélat d'activité physiologique associé au signal et donc une interprétation fonctionnelle de l'image. Le test T semble considéré comme un outil statistique du type « boîte noire » auquel on confie le soin de traiter le risque. On peut également émettre l'hypothèse que les élèves ont développé des connaissances de manière empirique. Ainsi, la manipulation de l'image, le feed-back du logiciel permettant de voir l'ampleur des zones colorées diminuer au fur et à mesure que la borne statistique inférieure augmente, semblent permettre de mieux comprendre la signification (ou la fonction) du test T.

L'absence de référence à la notion de test statistique pour résoudre le problème lors de la première séance ne permet pas de se prononcer sur la compréhension de la notion de test statistique par les élèves des six autres binômes, ni d'établir de lien avec leur paramétrage des images. Parmi les deux binômes qui avaient répondu correctement à la question plus technique d'utilisation du test T à la séance précédente, l'un (5B) a appliqué des bornes statistiques suffisamment significatives pour repérer les deux zones cérébrales recherchées mais l'autre non (2B).

3. Analyse et commentaires

3.1. Construction de la notion de test statistique

Lors de notre expérimentation, la notion de test statistique est apparue difficile à construire, parce qu'elle correspond à un mode de réflexion peu usité en classe de SVT ou même de mathématiques, mais aussi parce qu'elle est peu connue des enseignants de SVT eux-mêmes, ainsi que de leurs collègues de mathématiques. Le nombre d'élèves observés dans cette étude empirique est faible et nos résultats mériteraient d'être confirmés par de nouvelles expérimentations, mais ils apportent des pistes concernant les situations à mettre en œuvre pour aborder cette notion. Impliquer les élèves, par un problème qui les touche et dont la solution n'est pas évidente en regardant la série de données, paraît être une voie envisageable. Dans la situation d'apprentissage que nous avons mise en place, la question, souvent discutée entre adultes et adolescents, de l'effet de la musique sur la concentration, l'aspect ludique de la mesure, et le travail de chacun sur ses propres performances, ont permis une motivation importante des élèves et une forte dévolution du problème. Le logiciel *Réaction*, qui permet d'obtenir rapidement de nombreuses mesures et de mettre en œuvre facilement des protocoles de mesures comparatifs, s'est révélé être un outil adapté à ce type de problématique. Utilisé dans le cadre de l'étude de l'influence d'un fond sonore sur le temps de réaction, il a permis d'obtenir des résultats très différents d'un élève à l'autre, et de poser la question « Comment décider si la différence est significative ? », comme le montre l'extrait d'un rapport d'expérimentation rédigé par l'enseignant qui a renouvelé cette séance l'année suivante : « *Les moyennes ont été inscrites au tableau par groupe et nous avons commenté. La musique douce avait-elle une influence sur le temps de réaction ? Le panel de résultats était suffisant pour que la discussion soit rapidement intéressante.*

Un groupe avec 2 moyennes quasi identiques (1 ms d'écart sur la moyenne), un groupe avec 150 ms d'écart et les autres entre les deux. Les élèves ont donc admis que pour le premier groupe elle n'avait pas d'influence et que pour le deuxième elle en avait. Se posait alors la question des autres groupes. Je leur ai demandé leur avis d'abord, et sur quoi ils le fondaient, puis d'indiquer à partir de quel écart ils considéraient que la différence était significative. Spontanément sont sortis des pourcentages (20 %, 30 %) ». La discussion des résultats à l'échelle de la classe n'a pas été abordée avec les élèves durant cette expérimentation, mais elle mériterait également d'être explorée, en lien avec la problématique d'étude à l'échelle d'un groupe d'individus ou d'une population. L'étude des données au niveau de la classe tend à montrer un effet de la musique moins important que les dispersions des temps de réaction entre individus, mais également que la variance individuelle des temps de réaction est relativement importante.

La possibilité de manipuler le modèle mathématique représenté par le test de Student, à la fois sous forme d'une courbe, mais aussi sous forme de cellules de tableur permettant de faire correspondre à une statistique de test un risque de se tromper et réciproquement, nous paraît aussi être un élément important de la situation. La comparaison de plusieurs séries de données, ou la modification de certaines données pour en mesurer l'effet sur la statistique de test, pourraient d'ailleurs être également des pistes envisageables, sans doute avec des objectifs d'apprentissage plus mathématiques. Dans la situation proposée, le test de Student reste en effet une boîte noire : les bases théoriques qui justifient la validité de ce test ne sont pas abordées. Mais la boîte noire se situe à un degré de granularité fin : les élèves utilisent un modèle mathématique, savent que c'est un modèle mathématique, connaissent les données utilisées et la forme de la réponse apportée. Le choix du risque encouru relève de leur décision. Le modèle permet donc d'apporter une réponse nuancée – et non pas binaire – à la question de la significativité de la différence dans les mesures effectuées. De ce point de vue, notre étude met en évidence l'importance de se situer dans un seul registre, celui du risque par exemple, pour ne pas perdre les élèves en passant du registre du risque à celui de la confiance.

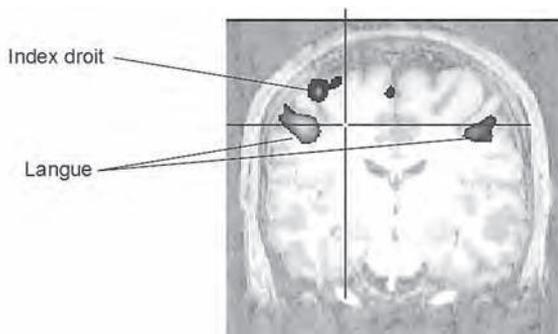
3.2. Mobilisation de la notion de test statistique pour le paramétrage des IRMf

Pour explorer les liens réalisés par les élèves entre les deux séances proposées, les échanges relatifs au réglage des bornes d'affichages et les références à la première séance, retrouvés dans deux des quatre clavardages disponibles ont été analysés. Il s'agit d'échanges entre les binômes 1B-2B et 7B-8B. On peut remarquer dans les deux cas que les deux binômes associés avaient des productions différentes avant la collaboration : l'un ayant réussi à localiser les zones cérébrales recherchées et l'autre non. Les élèves font référence aux bornes en utilisant le mot « curseur », qui correspond à l'aspect fonctionnel de l'interface et la possibilité de modifier ces bornes, et les termes « Inf » et « Sup », qui sont écrits sur l'interface.

Les binômes 7B et 8B n'avaient pas mobilisé la notion de test statistique pour résoudre leur problème lors de la première séance. Leur clavardage (figure 3) peut être qualifié de « disputational talk », au sens de Mercer (1996). En effet, on observe une opposition entre eux et le refus de 8B de revenir sur son paramétrage. La collaboration n'a pas lieu. Les bornes statistiques ayant été indiquées sur la diapositive construite par 7B (figure 4), sous la forme « Inf : 56 Sup : 100 » et « Inf : 69 Sup : 100 », 8B demande à 7B « *Explique pourquoi l'écart entre tes curseurs est si étroit ?* » (12:07:12). L'argumentation développée par 7B ne fait pas intervenir la notion de test statistique, de risque d'erreur, ou de spécificité. Elle exprime une approche visuelle du problème « *ben c'est pas très clair* » (12:11:37), « *et nous c'est beaucoup plus clair* » (12:12:47), ou bien fait référence à l'enseignant pour avancer un argument d'autorité « *ben non, c'est le prof qui vient de nous le dire* » (12:12:29), sans toutefois reprendre les arguments abordés dans la discussion avec l'enseignant. L'assertion « *pour que ce soit plus précis* » (12:11:37), nous semble aussi faire plutôt référence à une précision visuelle (la taille des taches de couleur observées), qu'à une éventuelle précision mathématique. On peut penser que le binôme 7B a fait le lien entre l'étendue des régions colorées et l'intervalle entre les deux bornes, sans doute en manipulant la borne inférieure. Il est possible que plutôt que de fixer un seuil statistique élevé, pour ensuite rechercher la zone cérébrale colorée spécifique de la stimulation tactile étudiée, les élèves 7B aient choisi les bornes pour obtenir une image conforme à leurs attentes, c'est-à-dire au contrat pédagogique (des zones cérébrales différenciées et précisément localisées). Le fait que les bornes inférieures soient différentes pour l'IRMf de stimulation de l'index et de la langue est cohérent avec cette hypothèse. S'ils avaient véritablement choisi une valeur de seuil inférieur, ils l'auraient vraisemblablement appliquée aux deux images de la même façon. On peut donc penser que, même si l'objectif d'obtention d'une image spécifique est atteint, la signification statistique de l'image n'a pas été comprise (ou construite) par le binôme 7B. La traduction sémiotique de valeur statistique en image colorée semble représenter un obstacle à la compréhension. Les élèves focalisent leur attention sur la représentation des résultats de l'étude (les couleurs) et perdent de vue la signification de ces résultats d'un point de vue statistique. Ils ne parviennent donc pas à effectuer une analyse en profondeur des codes et symboles affichés sur l'interface. Ainsi, les arguments de 7B ne réussissent pas à convaincre 8B, pour qui le choix des bornes a un caractère personnel « *chacun son choix* » (12 :12 :22). S'agit-il dans ce cas d'un choix orienté par la dimension esthétique de l'image ? Durant le clavardage, le binôme 8B modifie cependant ses diapositives, en ajoutant l'indication des bornes choisies à la façon de 7B, ayant ainsi repéré que cette information est importante, même si sa signification n'est pas maîtrisée. La manipulation du logiciel n'a donc pas permis à ces deux binômes de construire une représentation de l'IRMf en temps que carte statistique.

12:07:12	8B	<i>Explique pourquoi l'écart entre tes curseurs est si étroit ?</i>
12:07:30	7B	<i>Pour mieux voir.</i>
12:07:55	8B	<i>ok</i>
12:09:08	7B	<i>C'est quoi vos curseurs à vous ?</i>
12:09:48	8B	<i>c'est inf 30 et sup 100</i>
12:11:37	7B	<i>Bin c'est pas très clair, il fallait être plus haut là où vous avez mis 30, pour que ce soit plus précis.</i>
12:12:12	8B	<i>non c'est vous qui avez pris trop haut</i>
12:12:22	8B	<i>chacun son choix</i>
12:12:29	7B	<i>Bin non, c'est le prof' qui vient de nous le dire.</i>
12:12:47	7B	<i>et nous c'est beaucoup plus clair !!!</i>
12:13:03	8B	<i>et bah on s'en fou</i>
12:13:18	7B	<i>Bin recommencez !</i>
12:13:54	8B	<i>Non.</i>
12:14:06	7B	<i>Ok.</i>
12:14:27		<i>Échec de l'envoi du fichier « leboss_x_742752540189.xml » à 8B.</i>
12:15:45	7B	<i>Refaites en augmentant le curseur inférieur !!!!</i>
12:16:10	8B	<i>nn c tro long</i>
12:16:22	7B	<i>=/Feignasse !</i>
12:16:23	8B	<i>et puis c'était notre choix de prendre ces valeurs</i>

Fig. 3 : extrait du clavardage entre les binômes 7B et 8B.



Coupe coronale

Index droit : Inf : 56 Sup: 100
 Langue : Inf : 69 Sup : 100

Fig. 4 : diapositive produite par le binôme 7B.

Le binôme 1B qui avait compris la fonction du test statistique lors de la première séance collaborait avec le binôme 2B, qui lui n'avait pas mobilisé le résultat du test pour résoudre son problème, mais avait été capable de répondre correctement à la question technique correspondante (tableau 2). La collaboration est effective durant leur clavardage (figure 5) qui présente les caractéristiques de « *exploratory talk* » (discussion ou échange exploratoire) (Mercer, 1996). En effet, des éléments sont proposés dans le but de résoudre ensemble le problème ; ils sont critiqués et modifiés. Le raisonnement est visible. Alors qu'aucun des deux binômes n'avait indiqué sur ses diapositives les valeurs choisies pour les bornes statistiques, 1B questionne 2B sur ces valeurs dès le début du clavardage, en utilisant les termes « *contraste* », puis « *curseur* » (12:04:26 - 12:04:57). La signification statistique des bornes est comprise par 1B qui remarque que le choix fait par 2B « *laisse pas mal de marge d'erreur* » (12:09:02). On peut donc penser que la notion de test statistique travaillée la semaine précédente est ici réinvestie pour effectuer le paramétrage des IRMf. Le binôme 2B accepte de choisir les valeurs de 1B (12:10:46), mais se heurte au problème de ne plus voir aucune zone colorée sur son écran (12:11:31). En effet, la zone cérébrale recherchée par 2B n'est pas dans le plan de coupe qui s'affiche par défaut à l'ouverture de l'image. Pour la voir, il faudrait explorer le volume cérébral, ce que les élèves 2B n'ont pas su faire. Débute alors une négociation sur la valeur de la borne inférieure, qui doit permettre de visualiser une zone, tout en étant la plus élevée possible (12:12:16). Il est intéressant de constater que 2B fait à son tour référence à la séance précédente, et se questionne sur la relation précise entre les valeurs affichées sur l'interface et le risque pris (12:13:44). Les élèves 2B voudraient même connaître la façon de calculer le risque à partir des valeurs (12:15:17), ce qui correspond à l'activité réalisée lors de la séance précédente. Dans le cas de ces deux binômes, les notions de statistique travaillées lors de la séance précédente sont donc mobilisées et deviennent des outils pour la compréhension de la signification des IRMf lors de leur paramétrage. On peut d'ailleurs regretter que le lien mathématique entre la statistique de test et la valeur de la borne statistique inférieure ne soit pas explicite dans l'utilisation de l'interface.

Nous pouvons donc retenir que, dans ce cas, des élèves qui s'étaient approprié la notion de test statistique lors de la première séance, l'ont réinvestie lors du paramétrage des IRM, comme un outil qu'ils utilisent pour argumenter au cours des échanges entre binômes. Les binômes qui ne s'étaient pas approprié cet outil, ont choisi des stratégies de paramétrage des images basées sur des critères « visuels », revendiqués comme subjectifs, ou appliqués en référence à l'autorité de l'enseignant. Sans l'outil d'analyse statistique qu'est le test T, la construction de l'image reste donc une boîte noire et sa dimension statistique ne peut pas être appréhendée par les élèves. Une maîtrise minimale de certains concepts de la statistique est donc nécessaire pour comprendre la neuro-imagerie. Cependant, le transfert d'un modèle développé dans un registre graphique et numérique (le travail sur le test de Student lors de la séance 1) dans un registre sémiotique pictural

12:04:26	1B	<i>vous avez pris combien de contraste ?</i>
12:04:57	1B	<i>au niveau du curseur du haut</i>
12:05:02	2B	<i>deux</i>
12:05:09	1B	<i>lol</i>
12:05:16	1B	<i>ou pas</i>
12:05:16	2B	<i>atten</i>
12:05:40	2B	<i>et vous c combien</i>
12:05:41	1B	<i>nous on a pris 70 et 100</i>
12:05:59	1B	<i>'fin 70 en haut et 70 en bas 100 [...]</i>
12:08:25	2B	<i>32 en haut</i>
12:08:31	2B	<i>100 en bas</i>
12:09:02	1B	<i>ok, mais ca laisse pas mal de marge d'erreur</i>
12:09:47	2B	<i>ouai</i>
12:10:08	2B	<i>fau prendre le meme numero</i>
12:10:12	1B	<i>avec 70 et 100 c'est moins important</i>
12:10:20	1B	<i>important*</i>
12:10:46	2B	<i>on pren 70</i>
12:10:52	1B	<i>ok</i>
12:11:21	1B	<i>mais là on doit faire la carte à deux non?</i>
12:11:31	2B	<i>on peu pas prendre 70 il y a rien</i>
12:11:31	2B	<i>rien</i>
12:11:39	2B	<i>oui mai avec la notre aussi</i>
12:11:49	2B	<i>donc il fo prendre les meme donné</i>
12:12:16	1B	<i>oui, on prend moins alors mais faut pas avoir de trop basse valeur</i>
12:12:38	1B	<i>on entend tout</i>
12:13:33	1B]	<i>oui c'est important , mais on peut encore descendre; essayez avec 60 et 50</i>
12:13:44	2B	<i>si on pren 32 100, se sera combien notre truc de %</i>
12:14:47	1B	<i>bah le monsieur m'a dit qu'avec 70 on avait a peu près 70 % de juste</i>
12:15:09	1B	<i>donc a 30 y doit y avoir 70 % d'erreur</i>
12:15:17	2B	<i>mai comment on fai le calcule</i>
12:15:27	2B	<i>pour savoir</i>

Fig. 5 : extrait du clavardage entre les binômes 1B et 2B.

(les couleurs, la taille, la position dans l'espace des régions sur l'IRM construite durant la séance 2) est difficile et doit être accompagné. Clément (1996) avait proposé un travail de mosaïque dès les petites classes, pour préparer l'accès à la pixellisation des images numériques. La possibilité offerte par le logiciel EduAnatomist de manipuler les paramètres du modèle statistique en visualisant les conséquences sur

l'IRM fonctionnelle étudiée est également un moyen d'expérimenter ces transitions de registre, pour mieux comprendre la signification réelle de l'image scientifique. Les éléments de l'interface du logiciel qui permettent à l'élève de manipuler l'image doivent donc tenir compte à la fois des difficultés mises en évidence chez les élèves, et des relations existant entre les différents registres (statistique de test, couleur, taille, position dans l'image). Nos résultats ont d'ailleurs permis de proposer des modifications de l'interface du logiciel (Monod-Ansaldi *et al.*, 2011).

Les modèles statistiques ne sont pas les seuls outils à mettre en œuvre pour faciliter la compréhension de la signification des IRM fonctionnelles par les élèves. La prise en charge dans les situations d'apprentissage de la fonction de communication de ces images scientifiques nous paraît également cruciale ainsi qu'un travail sur les idéologies et les valeurs qu'elles peuvent mettre en jeu. Ainsi, l'approche avec les élèves de questions socialement vives mettant en jeux des IRM fonctionnelles, telles que les détecteurs de mensonges, ou le neuro-marketing, pourrait également permettre d'aborder la complexité de ces images dans leur aspect éthique et leurs usages sociaux. Il s'agirait alors de relier le domaine de validité de ces images avec leurs utilisations sociales : ce qu'elles disent et ce qu'on leur fait dire. De ce point de vue, il serait pertinent de travailler avec les élèves sur la variabilité interindividuelle qui caractérise le fonctionnement cérébral.

3.3. Traitement des données issues de mesures en classe de SVT

Cette étude nous permet également de proposer des pistes de réflexion pour l'exploitation des résultats obtenus lors d'activités expérimentales conduites en classe de biologie. Pour éviter un positionnement dogmatique ou inductif, qui consiste à conclure rapidement à partir de deux mesures sans se questionner sur la significativité des données obtenues, il paraît envisageable d'effectuer un plus grand nombre de mesures, permettant d'acquérir plus de données, et d'utiliser des outils statistiques plus fréquemment. Le difficile test de Student, utilisé ici en référence aux protocoles de construction des IRMf n'est pas toujours nécessaire. La comparaison des intervalles de confiance de deux séries de mesures (Schwartz, 2010), traités en mathématiques en classe de seconde⁵, est aussi une approche envisageable. Si les intervalles sont disjoints, on conclut que la différence est significative. La conduite d'investigations scientifiques en classe peut être l'occasion d'aborder avec les élèves cette question méthodologique, parfois même dans un cadre pluridisciplinaire, par exemple dans le cadre de l'enseignement exploratoire Méthodes et pratiques scientifiques (MPS), qui articulent des enseignements de mathématiques, de SPC et de SVT. Une telle approche permettrait de s'abstraire d'une vision des mathématiques comme « discipline outil » des sciences expérimentales, la situation problème de biologie donnant sens aux notions mathématiques abordées. Des travaux sont actuellement

⁵ Bulletin officiel du ministère de l'Éducation nationale, n° 30, 23 juillet 2009.

en cours pour explorer cette voie. Le développement de telles approches ne pourra faire l'économie d'un accompagnement spécifique des enseignants qui, à ce jour, ne sont pas suffisamment formés à ces outils et ces démarches.

Conclusion

La mobilisation d'outils mathématiques et statistiques en sciences de la vie et de la Terre n'est pas spécifique du champ des neurosciences. Nombre de phénomènes biologiques sont ainsi modélisés avec un fort degré de formalisation. Les méthodes probabilistes et les outils statistiques constituent une des bases de la biologie actuelle (Girault & Girault, 2004), s'insérant dans des approches systémiques, associant recueil et traitement de données à des démarches de modélisation, quand l'expérimentateur est confronté à des causalités plurifactorielles ou à des facteurs en interaction, pour lesquels il doit renoncer à tout maîtriser (Coquidé, Fortin & Rumelhard, 2009). Dans les classes, la conduite de démarches expérimentales sous une forme caricaturale prédomine. Introduire plus fréquemment une approche probabiliste et modélisante dans les activités des élèves, notamment à l'occasion d'investigations, est un enjeu actuel en lien avec l'épistémologie scolaire. Une piste possible pourrait consister à aborder des problèmes complexes, par un travail co-disciplinaire, aboutissant à la co-construction de sens à propos d'un même objet d'étude (Blanchard-Laville, 2000), en respectant et articulant les positions épistémologiques de chacune des disciplines (Chevallard, 2004). Toutefois, dans l'approche statistique au niveau des populations en biologie, on doit garder à l'esprit la confusion possible entre norme et moyenne, entre normes scientifiques et normes sociales.

Réjane MONOD-ANSALDI

rejane.monod-ansaldi@ens-lyon.fr

Grégoire MOLINATTI

gregoire.molinatti@univ-montp2.fr

Valérie FONTANIEU

valerie.fontanieu@ens-lyon.fr

Daniel DEVALLOIS

daniel.devallois@ac-grenoble.fr

Éric SANCHEZ

eric.sanchez@ens-lyon.fr

Remerciements

Nous remercions Claudine Schwartz pour ses retours sur l'aspect statistique de ce travail, et l'ensemble des enseignants qui ont travaillé sur le logiciel Réaction, le logiciel *EduAnatomist* et la banque *NeuroPeda*. Merci également au lycée Madame-de-Staël de Saint-Julien-en-Genevois qui a accueilli cette expérimentation et aux élèves de première S qui s'y sont prêtés.

BIBLIOGRAPHIE

- BLANCHARD-LAVILLE C. (2000). De la co-disciplinarité en sciences de l'éducation, *Revue française de pédagogie*, n° 132, p. 55-66.
- BROUSSEAU G. (1998). *Théorie des situations didactiques*. Grenoble : La Pensée sauvage.
- BUNGE M. (1975). *Philosophie de la physique*. Paris : Seuil.
- CHEVALLARD Y. (2004). *Vers une didactique de la codisciplinarité. Notes sur une nouvelle épistémologie scolaire*. Communication aux Journées de didactique comparée, Lyon, 3-4 mai 2004. Version du 19 mai 2004. En ligne : <<http://yves.chevallard.free.fr/spip/spip/recherche.php3?recherche=codisciplinarit%E9>>. (Consulté le 13 septembre 2011).
- CLÉMENT P. (1994). La difficile évolution des conceptions sur les rapports entre cerveau, idées et âme. In A. Giordan, Y. Girault & P. Clément (éd.), *Représentation, conceptions, connaissances*, Berne : Peter Lang, p. 73-90.
- CLÉMENT P. (1996). L'imagerie biomédicale : définition d'une typologie et proposition d'activités pédagogiques. *Aster*, n° 22, « Images et activités scientifiques », p. 87-126.
- CLÉMENT P. (1997). Cerveaux d'hommes et de femmes : l'idéologie était déjà dans la revue Nature. In A. Giordan, J.L. Martinand & D. Raichvarg (éd). *Actes Journées internationales sur l'éducation scientifique*. Chamonix : université Paris Sud, p. 267-272.
- CLÉMENT P. (2001). Using complex images in the popularization of science: Scientific or ideological agenda? in J.F. Rouet, J. Levonen & A. Biardeau (éd.), *Multimedia learning: cognitive and instructional issues*, Londres : Pergamon, p. 87-98 et p. 182-183.
- CLÉMENT P. & FISSEUX C. (1999). Opacity of radiography, perplexity of teachers and pupils in primary school. In M. Bandiera, *Research in Science Education in Europe*, Dordrecht : Kluwer, p. 15-21.
- COQUIDÉ M., FORTIN M. & RUMELHARD G. (2009). L'investigation : fondements et démarches, intérêts et limites. *Aster*, n° 49, p. 51-78.
- FONTANIEU V., DEVALLOIS D., MOLINATTI G. & MONOD-ANSALDI R. (2010). Des temps de réaction. En ligne sur *Statistix*, <<http://www.statistix.fr/spip.php?article78>>. (Consulté le 14 septembre 2011).
- FORISSIER T. & CLÉMENT P. (2000). L'identité biologique n'est pas que génétique : un défi pour un enseignement citoyen. *Symposium BioEd: The Challenge of the Next Century*, Paris, 15-18 mai 2000.
- GIRAULT Y. & GIRAULT M. (2004). *L'aléatoire et le vivant*, Québec : Presses de l'université Laval.
- GOULD S. J. (1983). *La mal-mesure de l'homme*. Paris : Ramsay.

- HERGUETA S., LEMIRE F. & GIRAULT Y. (1997). *Résultats et analyse préliminaire des études évaluatives menées sur le projet d'exposition « cerveaux » par l'équipe « médiation muséale » de la Grande galerie de l'Évolution. Questionnaire « Attentes et représentations »*. Paris : Muséum national d'histoire naturelle [non publié].
- JACOBI D. (1987). *Textes et images de la vulgarisation scientifique*. Berne : Peter Lang.
- LATOUR B. (1985). Culture Technique : une introduction à l'anthropologie des sciences et des techniques. *Les vues de l'esprit*, n°14, p. 4-30.
- LE MAREC J., BOUCHER J. & HIARD S. (1992). *L'homme et la santé. Évaluation préalable, rapport interne*. Paris, Cité des sciences et de l'industrie, cellule évaluation, direction des expositions.
- MAFFEO V. (2007). Plasticité cérébrale et enseignement, de l'utilisation des images médicales. In M. Coquidé & S. Tirard (éd.), *Neuroplasticité, enseigner de nouveaux savoirs ou un nouveau regard ?* Paris : Vuibert/adapt-snes, p. 105-128.
- MALABOU C. (2004). *Que faire de notre cerveau ?* Paris : Bayard.
- MEIN M.-T. & CLÉMENT P. (1988). Comment se représente-t-on aujourd'hui notre cerveau ? In A. Giordan, *JIES X - Communication, éducation et culture scientifique et industrielle*, ARThèque, ENS Cachan, STEF. En ligne : <<http://artheque.ens-cachan.fr/items/show/326>>. (Consulté le 13 avril 2012).
- MERCER N. (1996). The quality of talk in children's collaborative activity in classroom. *Learning and Instruction*, n° 6, p. 359-377.
- MOLINATTI G. (2007). *Médiation des sciences du cerveau. Approche didactique et communicationnelle de rencontres entre neuroscientifiques et lycéens*. Thèse de doctorat, Paris : Muséum d'histoire naturelle.
- MONOD-ANSALDI R., DAUBIAS P., RAMI G., SANCHEZ É., & MOLINATTI G. (2011). Évaluation didactique et ergonomique de l'EIAH EduAnatomist pour l'éducation à l'image scientifique. In M. Bétrancourt, C. Depover, V. Luengo, B. De Lièvre & G. Tamperman (éd.), *EIAH 2011 À la recherche de convergence entre les acteurs des EIAH*. Mons, Éd. de l'Umons , p. 329-340.
- SAVY C. (2005). *Comment des enfants de 5 à 11 ans dessinent ce qu'ils ont dans la tête ? Leurs conceptions retrouvent-elles les modèles historiques ?* Thèse de doctorat, université Lyon 1.
- SCHWARTZ C. (2010). *Lien entre test et intervalle de confiance : cas particulier d'une moyenne pour des échantillons gaussiens ou de grande taille*. En ligne : <http://acces.inrp.fr/acces/ressources/neurosciences/temps-de-reaction-investigation-variabilite-et-traitements-statistiques-des-donnees/comprendre-1/traitements-statistiques-de-series-de-donnees/remarques_test_et_IC.pdf>. (Consulté le 21 juin 2012).

- SEGBARTH C. & ZIEGLER A. (2004) *Introduction à la neuroimagerie fonctionnelle et métabolique par RMN*. Journées de formation, Grenoble, 21-26 novembre 2004.
- VAN BLYENBURGH N.H. (1989). *Une étude du décalage entre les connaissances du public et le savoir scientifique en biologie humaine : évaluation des conceptions, analyse des obstacles et réalisation d'aides didactiques*. Thèse de doctorat, faculté des sciences de l'université de Genève.
- VIDAL C. & BENOIT-BROWAEYS D. (2005). *Cerveau, sexe et pouvoir*. Paris : Belin.

Abstracts • Zusammenfassungen • Resúmenes

Statistical tests and cerebral MRI in a scientific class in Première (Year 11)

Functional magnetic resonance imaging (fMRI) has been instrumental in making a major breakthrough in understanding how the brain works. These images, which are far from simply being snapshots of the working brain, result from complex development protocols and have a statistical meaning. Understanding their statistical nature is difficult but necessary to gauge their validity and develop a critical attitude to their use, especially in the media. The pupils' representations of this type of images do not usually integrate this statistical dimension. We present and analyse an innovative learning sequence towards improving the analysis of these images. It is a modelling lesson including the setting of fMRI with the EduAnatomist software preceded by an activity to appropriate the notion of statistical test (test T). Our findings reveal the inherent difficulties when pupils have to take account of the statistical aspect of biological images. Didactical aids are suggested towards opening – at least partly – some of the black boxes that reflect the development of this type of image.

KEYWORDS • Neurosciences, nuclear magnetic resonance imaging, modelling, statistics, representation, problem-based learning.

Statistische Tests und Echoenzephalographie in der 11. Klasse mit Leistungskurs Mathe und Wissenschaft

Die Bilder von Echoenzephalographie haben einen bedeutenden Fortschritt im Verständnis der Funktionsweise des Gehirns ermöglicht. Viel mehr als einfache „Fotografien“ des funktionierenden Gehirns, ergeben sich diese Bilder aus komplexen Bildungsprotokollen und haben eine statistische Bedeutung. Das Verständnis ihres Wesens ist schwierig, aber notwendig, um ihren Gültigkeitsbereich zu erfassen und eine kritische Haltung gegenüber ihrer Benutzung - vor allem in den Medien - einzunehmen. Die von den Schülern entwickelten Vorstellungen dieser Bilder berücksichtigen in der Regel diese statistische Dimension nicht. Wir analysieren eine innovative Unterrichtssequenz zum besseren Verständnis dieser Bilderkategorie. Es handelt sich um eine Modellierungssequenz, die die Parametrierung von Bildern aus der Echoenzephalographie mit der Software EduAnatomist beinhaltet. Dieser Sequenz ging eine Aktivität zur Aneignung des Begriffs statistischer Test (T Test) voraus. Unsere Ergebnisse zeigen Schwierigkeiten, was die Berücksichtigung durch die Schüler von statistischen Aspekten der biologischen Bilder angeht. Wir stellen didaktische Denkanstöße vor, um einige „Blackboxen“ über die Konstruktion von dieser Bilderkategorie mindestens einen Spalt zu öffnen.

Test estadísticos e IRM cerebrales en clase de 1^{re} S

Las imágenes IRM funcionales (IRMf) han permitido un avance importante en la comprensión del funcionamiento cerebral. Mucho más que meras « fotografías » del cerebro en funcionamiento, estas imágenes son resultados de protocolos de construcción complejos y tienen una significación estadística. La comprensión de su naturaleza estadística es difícil, pero necesaria para comprender su dominio de validez y construir una actitud crítica frente a su uso, en particular en los medios. Las representaciones de estas imágenes construidas por los alumnos suelen no tomar en cuenta esta dimensión estadística. Presentamos y analizamos una secuencia de aprendizaje innovadora aplicada para favorecer la comprensión de este tipo de imagen. Se trata de una secuencia de modelización que comprende el parametraje de IRMf con la ayuda del programa EduAnatomist, precedida de una actividad de apropiación de la noción de test estadístico (test T). Nuestros resultados ponen de manifiesto las dificultades inherentes a la toma en cuenta por parte de los alumnos del aspecto estadístico de las imágenes biológicas. Proponemos pistas didácticas para abrir, por lo menos en parte, ciertas cajas negras que corresponden a la construcción de este tipo de imagen.