

*Centre National de la Recherche Scientifique*

# Rapport d'activité scientifique

Laurent U PERRINET

*Pour évaluation par les sections du Comité national*



Équipe NEural OPERations in TOpographies (NeOpTo)
Institut de Neurosciences de la Timone
UMR 7289, CNRS / Aix-Marseille Université
27, Bd. Jean Moulin, 13385 Marseille Cedex 5, France
<a href="https://laurentperrinet.github.io/">https://laurentperrinet.github.io/</a>
<a href="mailto:Laurent.Perrinet@univ-amu.fr">Laurent.Perrinet@univ-amu.fr</a>

16 Février 2022

# Table des matières

<b>1</b>	<b>Curriculum Vitæ</b>	<b>2</b>
1.1	Présentation rapide . . . . .	2
1.2	Diplômes & titres universitaires . . . . .	2
1.3	Expérience scientifique professionnelle . . . . .	3
<b>2</b>	<b>Résumé de mon activité scientifique</b>	<b>5</b>
<b>3</b>	<b>Rapport d'activité sur la période de décembre 2016 à décembre 2021</b>	<b>6</b>
3.1	Introduction : codage prédictif et code neural . . . . .	6
3.2	Activités de recherche dans l'équipe InViBe (2012-2018) . . . . .	6
3.2.1	Protocoles pour l'étude de la détection du mouvement [Vacher16] .	7
3.2.2	Caractérisation fonctionnelle de la détection du mouvement [Ravello19]	8
3.2.3	Rôle fonctionnel des interactions latérales dans l'intégration spatio-temporelle . . . . .	9
3.2.4	L'effet du flash retardé comme preuve pour le traitement prédictif dans les cartes topographiques [KhoeiMassonPerrinet17] . . . . .	10
3.3	Activités de recherche dans l'équipe NeOpTo (2018-2022) . . . . .	12
3.3.1	Modélisation de réseaux de neurones impulsionnels . . . . .	12
3.3.2	Corrélat neural du mouvement apparent . . . . .	13
3.3.3	Résumé . . . . .	14
3.4	Productions scientifiques sélectionnées . . . . .	15
<b>4</b>	<b>Enseignement, formation et diffusion de la culture scientifique</b>	<b>17</b>
4.1	Encadrement de thèse et post-doctorants . . . . .	17
4.2	Participation à des activités grand public . . . . .	17
4.3	Collaboration artistique . . . . .	17
4.4	Enseignement . . . . .	18
<b>5</b>	<b>Transfert technologique, relations industrielles et valorisation</b>	<b>18</b>
5.1	Contrats et collaborations . . . . .	18
5.2	Développements de logiciels . . . . .	18
5.2.1	Mouvements des yeux et mouvement . . . . .	18
5.2.2	Biologically-Inspired Computer Vision . . . . .	19
5.2.3	Promotion du logiciel libre . . . . .	19
5.3	Expertise scientifique . . . . .	19
<b>6</b>	<b>Encadrement, animation et management de la recherche</b>	<b>19</b>
<b>7</b>	<b>Objectifs / Projet de recherche de janvier 2022 à janvier 2027</b>	<b>20</b>
7.1	Motivation : Dynamique des calculs neuronaux sous-jacents au traitement visuel . . . . .	20
7.2	Problèmes ouverts des processus visuels prédictifs . . . . .	23
7.2.1	Les défis de la représentation de l'information visuelle dans les réseaux de neurones impulsionnel (SNN) . . . . .	23
7.2.2	Le rôle des vagues d'activité corticales dans le traitement dynamique de l'information visuelle . . . . .	24
7.2.3	Propriétés intégratives des aires corticales . . . . .	24
7.3	Résumé et conclusions . . . . .	26

# 1 Curriculum Vitæ

## 1.1 Présentation rapide

- 48 ans, né le 23 Février 1973 à Talence (Gironde, France).
- Directeur de Recherche (DR2, CNRS), Affiliation : Équipe NEural OPerations in TOpographies (NeOpTo) - Institut de Neurosciences de la Timone (UMR 7289, CNRS / Aix-Marseille Université)
  - Adresse : 27, Bd. Jean Moulin, 13385 Marseille Cedex 5, France
  - E-mail : <mailto:Laurent.Perrinet@univ-amu.fr>
  - Téléphone : 04 91 32 40 44
  - URL : <https://laurentperrinet.github.io/>

## Résumé quantitatif des contributions

- 51 publications dans les revues avec comité de lecture (dont 4 en révision, 16 en premier auteur, 15 en dernier auteur), avec un total de 2916 citations, indice h= 26 et indice i10= 46<sup>1</sup>,
- 120 publications dans des actes de congrès avec comité de lecture,
- 4 livres et 6 chapitres de livres,
- 53 conférences invitées (dont 19 dans des congrès internationaux),
- 3 thèses en cours de direction (Hugo Ladret, Jean-Nicolas Jérémie, Antoine Grimaldi) et 3 finalisées (Angelo Franciosini, Victor Boutin, Mina Khoei),
- 4 Post-docs dirigés (Wahiba Taouali, Nicole Voges, Alberto Vergani, 1 à venir),
- 4 thèses co-dirigées (A Gruel, Mansour Pour, JB Damasse, J. Kremkow),
- 1 contrat en PI local (150k€), 1 AAP (100k€), 3 bourses de thèse obtenues,
- 20 contrats en collaborateur (dont 12 ANRs).

## 1.2 Diplômes & titres universitaires

### Habilitation à Diriger des Recherches, AMU, Marseille

2017

École Doctorale Sciences de la Vie et de la Santé, Aix-Marseille Université, France.  
Sous le titre “Codage prédictif dans les transformations visuo-motrices”, j’ai défendu mon Habilitation à Diriger des Recherches le 14 avril 2017.

Le jury était constitué des Prof. Laurent Madelain (Université Lille III), Dr. Alain Destexhe (Université Paris XI, Rapporteur), Prof. Gustavo Deco (Universitat Pompeu Fabra, Barcelona, Rapporteur), Dr. Guillaume Masson (Aix-Marseille Université), Dr. Viktor Jirsa (Aix-Marseille Université, Rapporteur) et du Prof. Jean-Louis Mege (Aix-Marseille Université).

### Doctorat de Sciences cognitives ONERA/DTIM, Toulouse

1999-2003

Titre : *Comment déchiffrer le code impulsionnel de la Vision ? Étude du flux parallèle, asynchrone et épars dans le traitement visuel ultra-rapide*. Allocataire d’une bourse MENRT, accueil à l’ONERA/DTIM.

- Cette thèse a été initiée par les résultats de la collaboration pendant le stage de DEA. Elle a été dirigée par Manuel Samuelides (professeur à SUPAÉRO et chargé de recherche à l’ONERA/DTIM) et co-dirigée par Simon Thorpe (directeur de recherche au CERCo)
- Participation et présentations à de nombreux colloques internationaux (IJCNN99, NeuroColt00, CNS00, CNS01, LFTNC01, ESANN02, NSI02). Participation aux écoles d’été “EU Advanced Course in Computational Neuroscience” à Trieste (Italie) et “Telluride Neuromorphic Workshop” au Colorado (États-Unis).
- En parallèle, j’ai participé à l’organisation d’une conférence sur les réseaux de neurones dynamiques (DYN\*2000). Je me suis aussi impliqué dans des activités d’enseignement (initiation à la programmation sous Matlab et théorie de la

---

1. Au 16 Février 2022, cf. <https://scholar.google.com/citations?user=TVyUV38AAAAJ&hl=fr>

probabilité) pour des élèves de première et deuxième année d'école d'ingénieur (à SUPAÉRO et à l'ENSICA, Toulouse) et des travaux dirigés de robotique (Traitement de l'image et reconnaissance d'objets au laboratoire d'Informatique et d'Automatique de SUPAÉRO).

- La thèse de doctorat a été soutenue le 7 février 2003 à l'université Paul Sabatier avec la mention "Très honorable avec les félicitations du jury". Le jury était présidé par Michel Imbert (Prof. Université P. Sabatier) et constitué par Yves Burnod (Directeur de recherche à l'INSERM U483) et Jeanny Hérault (Professeur à l'INPG, Grenoble).

#### **DEA de Sciences cognitives**

**1998-1999**

Univ. Paris VII, P. Sabatier, EHESS, Polytechnique, mention TB. Allocataire d'une bourse de DEA.

- Assistant de recherche, ONERA/DTIM (Département de Traitement de l'Image et de Modélisation), Toulouse (stage de DEA). 3/1999-7/1999
- Étude de l'apprentissage de type Hebbien de réseaux de neurones basés sur un codage par rang.
- Application à la reconnaissance de textures visuelles.
- Assistant de recherche, USAFB (Rome, NY) / University of San Diego in California (États-Unis). Étude de l'apprentissage autonome dans un système complexe de type automate cellulaire. 7/1999-8/1999

#### **Diplôme d'ingénieur SUPAÉRO, Toulouse, France.**

**1993-1998**

Spécialisation dans le traitement du signal et de l'image et en particulier dans les techniques des réseaux de neurones artificiels.

- Projets individuels sur la perception visuelle, la reconnaissance de locuteur, la compression de la parole et sur la reconnaissance de caractères.
- Ingénieur ALCATEL, Vienne (Autriche). Département du *Voice Processing Systems*. Ce 'stage long' volontaire, intégré à une formation de SUPAÉRO sur les systèmes industriels, impliquait toutes les étapes de la conception d'un système de messagerie téléphonique de technologie élevée : conception, prototype, contrôle de qualité et étude marketing. 9/1995-6/1996
- Assistant de recherche, JET PROPULSION LABORATORY (NASA), Pasadena, Californie. Département des Sciences de la Terre, Laboratoire d'imagerie radar, INTERFÉROMÉTRIE RADAR SAR APPLIQUÉE À LA GÉOPHYSIQUE 4/1997-9/1997
  - Programmation d'un processus de traitement d'images radar interférométriques SAR comprenant des corrections géographiques, une série de filtrages et un traitement d'interférométrie.
  - Étude et programmation d'un générateur de carte topographique.
  - Traitement des images obtenues pour surveiller la déformation de la croûte terrestre. Étude des tremblements de terre de Landers (Californie) et de Gulan (Chine).
- Assistant de recherche, CERCo (CNRS, UMR5549), Toulouse (stage de fin d'études d'ingénieur). Développement d'un réseau de neurones asynchrone appliqué à la reconnaissance de caractères. 4/1998-9/1998
  - Programmation du code du réseau de neurones asynchrones.
  - Étude et utilisation des statistiques non-paramétriques pour l'utilisation d'un code basé sur le rang d'activation des neurones.
  - Implantation d'une nouvelle règle d'apprentissage du réseau de neurones.

### **1.3 Expérience scientifique professionnelle**

**Directeur de Recherche (DR2)** (section CID51), INT/CNRS, Marseille

**2020-...**

**Chargé de Recherche Classe Normale**, INT/CNRS, Marseille

**2019-2020**

Au 1er janvier 2019, j'ai intégré l'équipe NeOpTo de Frédéric Chavane (DR, CNRS). J'implémente les modèles prédictifs dans des architectures bio-mimétiques.

**Chargé de Recherche grade 1**, INT/CNRS, Marseille

*2012-2019*

Au 1er janvier 2012, notre équipe a intégré l'Institut de Neurosciences de la Timone (UMR 7289, CNRS / Aix-Marseille Université ) à Marseille (direction Guillaume Masson). J'ai approfondi les modèles en me concentrant sur un codage probabiliste distribué appliqué à la boucle sensori-motrice.

**Mission longue** Karl Friston's theoretical neurobiology group (The Wellcome Trust Centre for Neuroimaging, University College London, London, UK). Collaboration avec Karl Friston sur l'application de modèles d'énergie libre aux mouvements oculaires.

*10/2010-*

*01/2012*

**Chargé de Recherche grade 2** (section 7), INCM/CNRS, Marseille

*2004-2012*

Sous la conduite de Guillaume Masson à l'INCM à Marseille, j'ai étudié des modèles spatio-temporels d'inférence dans des scènes naturelles en application de la compréhension des mouvements oculaires.

**Post-doctorat**, Redwood Neuroscience Institute (RNI), États-Unis

*2004*

Sous la conduite de Bruno Olshausen, j'ai comparé des modèles standards d'apprentissage avec une méthode originale centrée sur les potentiels d'action. Notamment, j'ai développé une méthode générique évaluant l'importance des processus homéostatiques dans l'apprentissage non-supervisé, qui a conduit à une publication dans le journal Neural Computation (référence A20-[Perrinet10shl]). J'ai ensuite étendu ce modèle à l'apprentissage spatio-temporels dans des flux video.

## 2 Résumé de mon activité scientifique

Mon objectif de recherche est d'étudier l'hypothèse selon laquelle on peut comprendre les liens entre la structure neurale, notamment l'arrangement topographique des neurones entre eux et la nature du signal nerveux, et la fonction des systèmes sensoriels comme l'optimisation de leur adaptation aux statistiques des scènes naturelles par des processus de type prédictif.

Plus précisément, je vise à étendre la compréhension des facultés sensorielles et cognitives sous la forme de modèles de réseaux de neurones impulsionnels qui réalisent de façon efficace des algorithmes de perception visuelle. En effet, les brèves impulsions du potentiel de membrane se propageant au fil des neurones sont une caractéristique universelle des systèmes nerveux et permettent de construire des modèles efficaces du traitement dynamique de l'information. Dans un but fonctionnel, je désire notamment implanter dans ces modèles des stratégies d'inférence grâce à des mécanismes d'apprentissage auto-organisés fixant les relations spatio-temporelles entre les neurones. Dans une démarche applicative, nous envisageons la création de nouveaux types d'algorithmes basés sur ces recherches.

### Résumé des travaux antérieurs et de leur impact scientifique

Mes travaux de thèse dirigés par Simon Thorpe et Manuel Samuelides ont permis d'explorer dans un cadre mathématique de nouveaux paradigmes de codage neural de type impulsif pour des images statiques [Perrinet03ieee] et en particulier comment ceux-ci peuvent être appris [Perrinet10shl]. Ceux-ci ont été étendus en collaboration avec Guillaume Masson à des modèles d'inférence statistique appliqués aux mouvements des yeux et à la boucle perception action [Simoncini12]. Un cadre théorique précis a été élaboré, sous le patronage de Karl Friston, pour tenir compte de la présence de délais temporels [PerrinetAdamsFriston14]. Actuellement, en collaboration avec Frédéric Chavane, j'étends des modèles dynamiques du traitement dynamique de scènes visuelles en mouvement [KhoeiMassonPerrinet17] dans des réseaux hiérarchiques [BoutinFranciosiniChavaneRuffierPerrinet20; Franciosini21]. Nous avons récemment mis en évidence le rôle de la précision dans ce traitement d'information [Ladret22] et qui nous permet de construire de nouveaux algorithmes neuromorphiques de type Spiking Neural Networks [Grimaldi22pami]. Une contribution de ces travaux est enfin d'apporter un regard nouveau sur la structure des réseaux neuraux, notamment dans les aires corticales visuelles [Chavane22].

### Mots clés

Perception, vision, détection du mouvement. Calcul parallèle événementiel, émergence dans les systèmes complexes, codage neural, réseaux de neurones impulsif. Inférence Bayésienne, minimisation de l'énergie libre, statistiques des scènes naturelles.

## 3 Rapport d'activité sur la période de décembre 2016 à décembre 2021

### 3.1 Introduction : codage prédictif et code neural

Le but de mon activité de recherche est de déchiffrer le “code neural”, c’est-à-dire de révéler dans la structure dynamique de l’activité neurale des algorithmes fonctionnels de traitement de l’information. Plus particulièrement je m’intéresse à comprendre comment le système visuel peut exploiter les régularités statistiques des scènes naturelles pour traiter le flux sensoriel de façon la plus efficace possible. Ce thème de recherche s’intègre donc plus généralement au problème de notre compréhension entre la structure du système nerveux central et de sa fonction. À ce titre, l’étude de l’intégration spatio-temporelle de l’information sensorielle est primordiale. En effet, les neurones présentent des contraintes physiologiques qui font que l’information sensorielle est locale aux premiers étages de captation du signal, alors qu’elle doit devenir globale et unique au niveau de la réponse comportementale. De plus, au fur et à mesure qu’ils montent les voies sensorielles jusqu’à la réponse motrice, ces signaux subissent de nombreuses transformations et subissent différentes sources de bruit. Ces problèmes se révèlent de façon saillante dans le système oculomoteur : Devant une réponse visuelle, comme l’image d’un prédateur pour une proie (et inversement), il est primordial à la survie du “sujet voyant” de pouvoir orienter son regard de façon efficace vers l’objet d’intérêt et de programmer une réponse adaptée le plus rapidement possible.

Dans ce cadre, le système oculomoteur procure un excellent modèle pour mettre à jour des processus de codage prédictif dans le code neural. En particulier, quel codage neural est le plus efficace pour l’intégration de l’information ? Comment intégrer différentes sources d’information (montante, associative, descendante) ? Quel est le meilleur compromis entre rapidité de la réponse et sa précision ? Afin de répondre à ces questions fondamentales pour les neurosciences, nous allons les aborder suivant les niveaux d’études suggérés par **Marr**<sup>83</sup>**empty citation**. Ceux-ci structurent mes axes de recherche suivant différentes approches :

**Fonctionnelle** Quelles fonctions sont à la source de la perception visuelle du mouvement ?

**Algorithmique** Comment exploiter le parallélisme et la dynamique des réseaux neuraux de façon efficace par rapport à la représentation de l’information visuelle ?

**Computationnelle** Comment implanter ces algorithmes dans la circuiterie neurale ? Quels enseignements ces modèles nous donnent pour déchiffrer le code neural ?

Cette dichotomie est bien sûr *a priori* arbitraire et constitue plutôt une grille de lecture pour aborder ces problèmes complexes.

Dans ce rapport d’activité scientifique, je résumerai mon activité de recherche en suivant cette grille. Nous débuterons dans la section 3.2 les travaux de recherche effectués au sein de l’équipe INViBE dirigée par Guillaume Masson à l’INT et nous nous sommes focalisés sur les liens entre codage neural et dynamique de la réponse oculaire. Je conclurai dans la section 3.3 sur les travaux de recherche effectués depuis 2018 au sein de l’équipe NEOPTO dirigée par Frédéric Chavane à l’INT.

### 3.2 Activités de recherche dans l’équipe INViBe (2012-2018)

En 2012, suivant la restructuration des neurosciences sur les différents sites du CNRS à Marseille, le laboratoire INCM a intégré un nouveau site sur le campus de la faculté de médecine de la Timone, l’Institut de Neurosciences de la Timone (INT). Ce déménagement a pris place fin 2010 et a donné lieu à une restructuration des équipes arrivantes pour assurer la cohésion de l’ensemble. Notamment, l’équipe DyVA est devenue l’équipe “inference and visual behavior” (INViBE). Tout en gardant de nombreux points commun avec les thèmes et méthodes développés à l’INCM, la formation de cette nouvelle équipe a permis de redéfinir son champ d’action. Notamment, l’accent a été mis sur l’intégration des pistes de recherche développées individuellement dans l’équipe et je vais développer dans cette section les principales contributions que j’ai pu apporter.

Une étape importante dans la structuration du thème de recherche a été la recherche de nouvelles sources de financement et de nouvelles collaborations. En particulier, nous avons obtenu un financement important grâce au projet BrainScaleS (commission européenne, contrat numéro FP7-269921), qui nous a permis d’envisager l’élaboration de nouveaux types d’algorithmes basés sur ces recherches. Une autre étape importante a été la collaboration avec Karl Friston à l’University College de Londres qui a permis d’étendre la portée théorique des modèles probabilistes que nous utilisons. Cette collaboration a pris la forme d’une mission longue de 14 mois (d’octobre 2010 à février 2012) sous l’invitation de Karl Friston et a permis l’établissement de nombreuses collaborations dans Londres et nationalement (notamment Jim Bednar à Edinburgh). Ces différents facteurs ont contribué à la restructuration du projet de recherche durant cette période que je détaille ici.

En particulier, tout en gardant une lecture proche des niveaux d’étude de Marr, nous allons progressivement les dépasser pour mettre en avant les collaborations entre différents niveaux. Pour cela nous allons d’abord étudier une approche héritée de l’ingénierie des systèmes pour caractériser le système oculomoteur (Sec. 3.2.1), pour ensuite étudier le rôle fonctionnel des interactions latérales dans l’intégration spatio-temporelle, et en particulier le rôle du codage prédictif (Sec. 3.2.3).

### 3.2.1 Protocoles pour l’étude de la détection du mouvement [Vacher16]

Mesurer la vitesse et la direction d’un objet en translation est une étape computationnelle cruciale pour bouger nos yeux, nos mains dans l’environnement, attraper un objet ainsi que percevoir l’organisation de la scène visuelle et de ses éléments. Par exemple, alors que nous avons une bonne connaissance des mécanismes perceptifs et neuronaux de l’encodage et du décodage de l’information de direction ainsi que des algorithmes biologiquement plausibles utilisés dans différentes espèces, comment le cerveau traite et représente l’information de vitesse reste largement incompris. Des neurones sélectifs à la vitesse ont été identifiés à différents niveaux hiérarchiques des voies visuelles chez l’homme et le singe mais ne savons toujours pas précisément comment cette sélectivité est construite. Ceci explique l’absence de modèle consensuel sur cette question. Une hypothèse de travail est que ces mécanismes neuronaux, et leurs corrélats perceptifs, combinent de façon non-linéaire l’information locale de mouvement extraite à travers plusieurs filtres spatiotemporels, prenant avantage de la structure multi-échelle des images naturelles. De plus, l’organisation perceptive de la scène et de ses parties doivent être pris en compte pour une intégration contextuelle et dépendant de la tâche. Enfin, le code neural sous-jacent à la perception de la vitesse reste lui aussi largement mystérieux et donc nous sommes loin de comprendre comment l’information de vitesse est décodé pour contrôler des réponses (oculo-)motrices et des jugements perceptifs. Récemment, nous avons proposé que l’estimation de la vitesse est intrinsèquement un problème multi-échelle et dépendant de la tâche [Simoncini12]. Nous avons défini un nouveau type de stimulus visuel de mouvement, des textures dynamiques dont la phase est aléatoire. Ces stimuli possèdent plusieurs des propriétés statistiques des images naturelles [Sanz12; Vacher15nips].

Une étape fondamentale a été franchie en important dans nos méthodes une approche héritée de l’ingénierie des systèmes. En effet, il est usuel pour caractériser le système oculomoteur d’utiliser des stimuli visuels simples comme des points, lignes ou des réseaux et de varier les paramètres de ces stimuli (contraste, orientation, direction, vitesse) pour en déduire la réponse comportementale. L’avantage de cette méthode est clairement la simplicité des stimuli. Toutefois, celle-ci s’accompagne paradoxalement avec le désavantage de créer des stimuli pour lesquels l’information peut être distribuée à différents niveaux de complexité structurelle. Ainsi une ligne en mouvement apporte un signal simple de mouvement, mais inclut intrinsèquement aussi des informations de haut niveau, comme l’alignement des différentes informations locales de mouvement. Une approche inverse est d’utiliser des stimuli écologiques en utilisant cette fois-ci des scènes naturelles. Le désavantage de ces stimuli est cette fois-ci que la complexité du stimulus est trop grande alors que l’on ne contrôle pas le contenu informationnel. Une solution pour caractériser le système oculomoteur est plutôt de faire l’hypothèse qu’il infère le mouvement d’un objet à partir d’un modèle



interne de ce mouvement. En paramétrisant ce modèle, on peut générer grâce au modèle direct des stimuli qui seront optimaux pour caractériser le système — sous réserve des hypothèses formulées.

J’ai formalisé un tel modèle de génération de texture qui a ensuite été implanté pour l’étude de la détection du mouvement [Sanz12], les *Motion Clouds* (voir <https://neuralensemble.github.io/MotionClouds/>). Il s’appuie sur la formalisation la plus simple d’un détecteur élémentaire de mouvement, le modèle “Motion Energy” [Adelson85]. Ce même modèle peut être de la même façon considéré comme la solution du problème inverse au modèle de conservation de la luminosité qui est souvent utilisé en vision par ordinateur [Aubert00]. Nous avons ensuite formulé ce modèle sous la forme d’une texture à phases aléatoires [Galerne10] en paramétrisant des axes perceptifs saillants (vitesse, direction, orientation) ainsi que les largeurs de bande (variabilité) qui leur sont associées. On obtient alors des stimuli aux statistiques proches des images naturelles, avec un jeu de paramètres à contrôler et avec une implantation simple<sup>2</sup>. Grâce à ce type de stimuli, nous avons pu par exemple caractériser la réponse oculomotrice en fonction de la richesse du contenu fréquentiel. Cette étude, parue dans *Nature Neuroscience* (impact factor 16.7), nous a permis de dissocier les différents processus non-linéaires en jeu dans une tâche décisionnelle ou perceptive : L’estimation du mouvement est intrinsèquement un problème multi-échelle et dépendant de la tâche [Simoncini12] (voir Fig. 1). Ces Motion Clouds constituent une base pour l’intégration de différentes études aux niveaux de la modélisation (pour valider les résultats théoriques), et aux niveaux physiologiques et comportementaux.

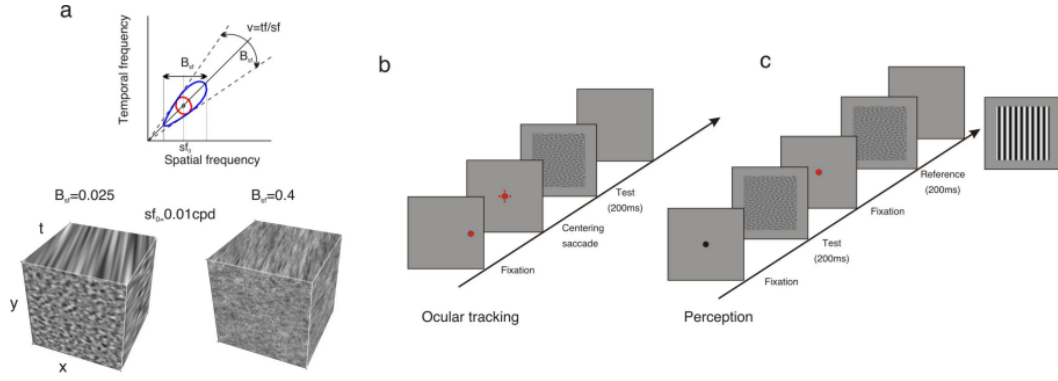
En particulier, nous avons développé autour de cet ensemble de stimuli différents axes de recherche. Dans une première étude, en collaboration avec Andrew Meso et Guillaume Masson, nous avons étudié l’estimation de la vitesse en fonction du contenu fréquentiel de la texture. Cette tâche est importante car elle nous permet de dissocier les contributions indépendantes des différents canaux dans la hiérarchie des voies visuelles et ainsi de caractériser finement la dynamique de l’intégration spatio-temporelle. Les résultats de psychophysique indiquent l’importance d’une information a priori telle que prédite par des modèles Bayésiens [Stocker06]. De façon plus surprenante, nous observons aussi des phénomènes de sur-estimation de la vitesse qui peuvent être expliqués en complétant ce dernier modèle [Meso13vss ; Meso14vss ; Vacher15nips]. Ces résultats expérimentaux peuvent être expliqués de façon globale en modélisant l’ensemble des transformations visuelles. En modélisant la génération aléatoire de textures en trois dimensions transformées par des opérations géométriques telles que des rotations, des zooms et des translations nous avons expliqué l’émergence d’a priori probabilistes qui nous ont permis de valider ces hypothèses de façon expérimentale. Ces travaux ont été publiés dans une conférence à au niveau scientifique [Vacher15nips] et sont aussi publiés dans une revue à comité de lecture [Vacher16].

### 3.2.2 Caractérisation fonctionnelle de la détection du mouvement [Ravello19]

Ce projet a réuni des psychophysiciens, des spécialistes du contrôle oculomoteur chez l’homme et des modélisateurs pour caractériser le système oculomoteur. Notre but est d’étendre le travail élaboré ensemble ces dernières années pour comprendre comment mouvements de poursuite et perception visuelle tirent avantage d’un traitement multi-échelle pour estimer le mouvement d’une cible. Nous poursuivons notre travail de conception mathématique de stimuli de haute dimensionnalité grâce à notre modèle génératif des images naturelles. Grâce à eux, nous recherchons comment la vitesse est encodé grâce à l’extraction de l’énergie de mouvement dans différents filtres spatiotemporels. En analysant les réponses motrices et perceptives, nous mettons ainsi en évidence les mécanismes non-linéaires (dépendance au contraste, superposition, supra-linéarité, précision...) sous-jacente à l’intégration des sorties de ces filtres et nous pouvons donc modéliser ces mécanismes dans une nouvelle version de notre modèle computationnel. Par exemple de tels stimuli ont récemment été utilisés sur la rétine de rongeurs [Ravello19]. De plus, nous avons testé

---

<sup>2</sup>. Le code de cet algorithme de génération de textures est disponible sur <https://github.com/NeuralEnsemble/MotionClouds>.



**FIGURE 1 – Construction de stimuli de type Motion Clouds pour différentes tâches comportementales.** Pour montrer que l’estimation du mouvement est intrinsèquement un problème multi-échelle et tâche dépendant, nous avons construit le protocole suivant. (a) Dans l’espace représentant la distribution spatio-temporelle de fréquence (espace de Fourier), chaque ligne passant par l’origine correspond à des stimuli se déplaçant à la même vitesse. Un simple réseau en translation pure correspond à un seul point dans cet espace. Nos stimuli de texture en mouvement ont une énergie distribuée dans une “ellipse” allongée le long d’une ligne de vitesse donnée, en gardant constantes les fréquences spatiales et temporelles moyennes. La bande passante spatio-temporelle a été manipulée par des largeurs de bandes différentes suivant ces axes, comme illustré par les cubes représentant les stimuli. La performance humaine a été mesurée pour deux tâches différentes, gérées en blocs parallèles. (b) Pour le suivi oculaire, les stimuli de mouvement ont été présentés pour une courte durée (200 ms) à la suite d’un centrage saccadique, visant à contrôler à la fois l’attention et la précision de la fixation. (c) Pour la discrimination de la vitesse, des stimuli de test et de référence ont été présentés successivement pour la même durée et les sujets ont été invités à indiquer si le stimulus de test a été perçue comme lente ou plus rapide que la référence. Les résultats présentés dans [Simoncini12] montrent alors des effets opposés de la largeur de bande dans ces deux tâches.

notre hypothèse que dans les scènes naturelles, ces mécanismes non-linéaires augmentent la précision des réponses et diminuent leur variabilité d’un essai à l’autre, ce qui conduit à des réponses motrices optimales. En comparant ces réponses motrices avec les jugements perceptifs, nous avons pu mettre en évidence une seconde hypothèse de travail : ces calculs non-linéaires dépendent de la tâche et du contexte sensoriel ou sensori-moteur [Simoncini12]. En particulier, nous avons vu dans quelle mesure les structures géométriques des scènes visuelles sont décisives pour la perception, au-delà du seul calcul de l’énergie de mouvement qui est utilisée par les mouvements oculaires.

### 3.2.3 Rôle fonctionnel des interactions latérales dans l’intégration spatio-temporelle

Les Motion Clouds ne sont qu’une étape dans la construction d’une approche systémique de la caractérisation du système oculomoteur. Ceux-ci sont en effet construits sur des hypothèses simples de synthèse du mouvement et vont nous servir de contrôle : Les scènes naturelles se caractérisent en effet par de nombreuses régularités statistiques qu’il faut alors introduire dans le modèle génératif de synthèse du mouvement. En particulier, il est plus probable que la trajectoire de l’image d’un objet suive une trajectoire continue plutôt que dis-continue. Il est remarquable de noter qu’alors les modèles qui considèrent une indépendance conditionnelle entre les mesures voisines de mouvements considèrent par conséquent qu’une trajectoire dis-continue est aussi probable qu’une trajectoire continue.

Partant de ce constat, j’ai proposé un modèle d’intégration spatio-temporelle qui propose d’inclure l’information a priori que le mouvement d’un mouvement est continu et doit donc être pris en compte dans la dynamique globale du modèle. Nous avons mis en évidence la

proximité de cette approche avec celle de **Burgi00empty citation** et aussi les limites de ce dernier modèle. En effet, établir des prédictions sur un espace de position et de vitesse entraîne une explosion du nombre combinatoire de prédictions possible. À l'inverse du modèle précédent qui considère une approximation grossière de l'espace topographique de position et de vitesse, nous avons utilisé une technique de traitement de l'image appliquée au suivi de contours, les filtres particuliers. J'ai implanté un tel algorithme qui fournit ainsi une plateforme de modélisation que nous appliquons à différents problèmes.

Une première application de cette méthode a consisté à résoudre le problème de l'ouverture. En effet, ce problème est remarquable car il souligne qu'une information locale (par exemple le mouvement d'une ligne infinie) n'est pas suffisant pour caractériser le mouvement global (le mouvement d'un segment fini dans une direction non perpendiculaire à son orientation). Classiquement, il est établi que ce problème est résolu par des mécanismes spécialisés qui détectent soit le mouvement au centre du segment, soit le mouvement des fins de lignes. Alors, il est courant d'admettre que cette dernière information résout le problème de l'ouverture par un processus de compétition. Grâce au modèle de codage prédictif basé sur le mouvement, nous avons au contraire montré que ces mécanismes spécialisés étaient plutôt une propriété émergente du système : il est suffisant pour résoudre les problèmes de l'ouverture [**Perrinet12pred**].

Il est étonnant de constater qu'avec des hypothèses simples —le codage prédictif basé sur le mouvement, nous pouvons ainsi caractériser des propriétés du système oculomoteur attribués classiquement à des mécanismes complexes et non-linéaires. Nous avons ainsi étendu, en collaboration avec Mina Khoei (en thèse FACETS-ITN) et Anna Montagnini (inViBe-INT), notre étude à un autre modèle classique pour l'oculomotricité : l'extrapolation du mouvement. Cette extension consiste à étudier le comportement du modèle lors d'une interruption transiente de l'entrée sensorielle. En effet, il est courant —par exemple lors d'un clignement de l'œil— que l'entrée sensorielle soit perturbée ou suspendue, et il est important pour le système de représenter une certaine continuité. Celle-ci se traduit à partir de certaines étapes dans la hiérarchie du système visuel par une activité neurale soutenue pendant l'interruption, comme montré dans le cortex infero-temporal chez le macaque [**Assad95**]. Nous avons mis en évidence que notre modèle pouvait répliquer un tel comportement et en particulier, nous avons apporté trois points : 1) une prédiction à la fois en position et en vitesse est nécessaire pour avoir un comportement robuste, 2) la représentation du mouvement perd progressivement de sa précision lors de l'interruption, et 3) le système doit avoir accumulé assez d'information pour être dans un mode de suivi. Nous avons publié ces résultats dans différentes conférences et journaux [**Khoei13jpp** ; **KhoeiMassonPerrinet17**]. Bien que ce modèle se base sur une conceptualisation de la propagation de l'information au sein d'une carte corticale (utilisant les filtres particuliers), nous verrons dans la section suivante (Sec. 7.2) qu'elle admet une implantation neuromimétique.

### 3.2.4 L'effet du flash retardé comme preuve pour le traitement prédictif dans les cartes topographiques [**KhoeiMassonPerrinet17**]

L'effet du flash retardé (Flash-Lag Effect, FLE) est une illusion visuelle populaire pour sa généralité et sa simplicité. Dans sa forme originale [**MacKay58**], on demande à l'observateur de fixer une croix centrale sur l'écran pendant qu'un point traverse l'écran avec un mouvement constant et horizontal. Lorsqu'il atteint le centre de l'écran, un autre point clignote brièvement juste en dessous du point en mouvement. Alors qu'ils sont parfaitement alignés verticalement, le point clignotant est perçu comme *en retard* par rapport au point en mouvement [**Perrinet19temps**]. Cette illusion visuelle a vu un regain d'intérêt scientifique avec le modèle d'extrapolation de mouvement [**Nijhawan02** ; **Nijhawan09**]. Cependant, d'autres modèles tels que le modèle de latence différentielle ou la post-diction ont également été proposés, de sorte que l'on ne savait pas encore clairement quel est le substrat neural du FLE. Ici, en étendant le modèle de compensation des délais [**PerrinetAdamsFriston14**], nous avons défini un modèle de traitement prédictif généralisé sur la topographie visuelle en utilisant une représentation interne du mouvement visuel [**Perrinet12pred**] qui utilise une diffusion anisotrope des informations 2-A.

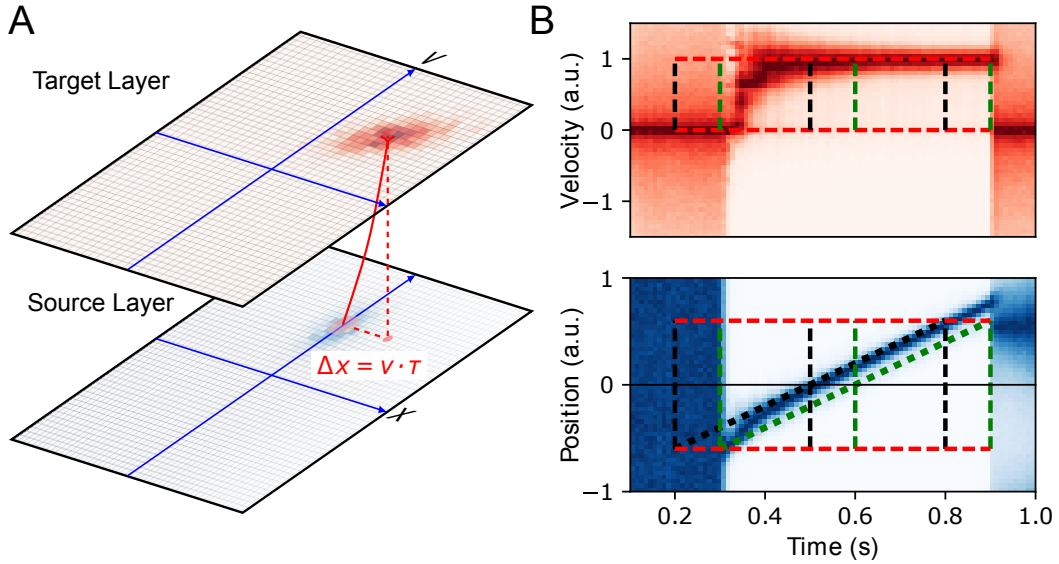


FIGURE 2 – Dans [KhoeiMassonPerrinet17], nous proposons un modèle de traitement prédictif sur une carte topographique. **(A)** le modèle consiste en une carte à deux couches : une cible source d'entrée intègre les informations des capteurs visuels. Par souci de simplicité, nous n'affichons ici que la dimension horizontale et cette carte représente sur chaque axe respectivement la position et la vitesse. En utilisant cette carte comme représentation de la croyance (ici en utilisant une fonction de distribution de probabilités), il est possible de projeter cette information sur une deuxième couche cible qui intègre l'information tout en compensant pour ce délai. Dans ce cas particulier, la vitesse est positive et donc l'information de position est transportée vers la droite. **(B)** Réponse d'un modèle compensant un délai de 100 ms pour le mouvement d'un point en mouvement. Représentation de la probabilité inférée de position et de vitesse avec compensation de délai en fonction des itérations du modèle (temps). Les couleurs plus foncées dénotent des probabilités plus élevées, tandis qu'une couleur claire correspond à une estimation improbable. En particulier, nous nous concentrons sur trois époques particulières le long de la trajectoire, correspondant à l'initiation, au milieu et à l'extinction. Le moment de ces trois époques est indiqué par des traits verticaux en pointillés : en noir pour le temps physique (réel) et en vert pour l'entrée retardée sachant un délai de 100 ms. Voir le texte pour une interprétation des résultats. (Reproduit de [KhoeiMassonPerrinet17] selon les termes de la Creative Commons Attribution License, © The Authors 2017.)

Le modèle que nous avons utilisé pour le FLE peut être utilisé avec n'importe quelle image. En particulier, un point lumineux isolé évoque une activité isotrope en expansion puis en contraction, tandis qu'un point en mouvement peut produire une onde de type soliton qui peut traverser une occlusion [Khoei13jpp]. Plus généralement, ce modèle peut être décrit comme utilisant le terme d'advection de l'équation de Navier Stokes qui gouverne la dynamique des fluides. Ainsi, les solutions à ces équations sont typiquement des "vagues" qui se déplacent sur la carte rétinotopique. Une caractéristique particulière de ces équations prédictives est qu'elles comprennent un terme d'amplification pour les mouvements rectilignes. Par conséquent, une fois qu'un objet commence à être suivi, sa position est prévue dans le futur, de sorte que la position et la vitesse sont mieux estimées. Au contraire, un point qui se déplace sur une trajectoire imprévisible est éliminée par le système. Ceci explique certains des comportements non linéaires, de type binaire, qui sont expliqués par ce modèle [Perrinet12pred]. Il est particulièrement intéressant à ce stade de comprendre si un tel modèle s'étend à d'autres stimuli ou si l'on peut préciser son implémentation neurale.

Appliquée à l'image du FLE, l'activité dans le modèle montre trois phases différentes, voir 2-B. Premièrement, il y a une augmentation rapide de la précision de la cible après la première apparition du point mobile (à 300 ms). En cohérence avec l'effet Fröhlich [Jancke10], le début de la trajectoire est vu "en avant" par rapport à sa position physique. Pendant la deuxième phase, le point en mouvement est suivi efficacement car sa vitesse et sa position sont correctement déduites. Comme observé dans l'illusion, l'extrapolation de mouvement prédit correctement la position actuelle et la position suit la position physique réelle du point (ligne pointillée noire) : Le point est perçu en avant de la trajectoire retardée du point (ligne pointillée verte). De façon équivalente, le flash est perçu en retard par rapport au point en mouvement. Enfin, la troisième phase correspond à la fin du mouvement. Le point en mouvement disparaît et l'activité correspondante disparaît dans la couche source à  $t = 900$  ms. Cependant, entre  $t = 800$  ms et  $t = 900$  ms, la position du point a été extrapolée et prédite avant la position finale. À  $t = 900$  ms, alors que l'information de mouvement est absente, l'information de position est encore transitoirement cohérente et extrapolée à l'aide d'une large distribution de vitesses centrée au préalable : Bien qu'elle soit moins précise, cette position du point à la fin du flash n'est donc pas perçue comme étant en avant du flash, mais *rétrospectivement* dû à cette nouvelle information.

### 3.3 Activités de recherche dans l'équipe NeOpTo (2018-2022)

Afin de confronter de tels modèles avec des données physiologiques et comportementales, j'ai rejoint l'équipe NeOpto en 2018. Nous allons résumer les travaux effectués dans cette période et tout d'abord montrer des modèles de réseaux neuraux impulsionnels à grande échelle, tout en formalisant une théorie de décodage de cette activité neurale en terme d'information visuelle (Sec. 3.3.1).

#### 3.3.1 Modélisation de réseaux de neurones impulsionnels

L'étude que nous menons sur les régularités statistiques —dans les trajectoires des objets et dans l'espace— se doivent d'être validés par les résultats expérimentaux. Pour ce faire, nous utilisons deux approches complémentaires. La première consiste à utiliser des simulations à grande échelle des conductivités que nous avons mis en évidence pour comprendre si ces principes s'étendent tels quels à des réseaux de neurones. La deuxième approche consiste à utiliser les données neurales collectées au laboratoire et d'utiliser nos modèles pour décoder de l'activité neurale l'information visuelle pertinente.

Dans le premier axe, nous avons implanté en collaboration avec Bernhard Kaplan, Anders Lansner (KTH, Suède) et Frédéric Chavane (inViBe-INT), et dans le cadre de BrainScaleS, des simulations à grande échelle d'une aire corticale implantant le codage prédictif basé sur le mouvement. Cette simulation est basée sur le savoir-faire du KTH en la matière et nous a permis de valider le modèle probabiliste à l'échelle neuro-morphique. Nous avons utilisé comme contrôle le protocole d'extrapolation du mouvement (voir plus haut et [Khoei13jpp]). Les résultats ont montré des résultats similaires aux résultats

théoriques (notamment les trois points évoqués plus haut), ainsi qu’une propriété reliée à l’implantation neurale. Durant l’interruption, la représentation probabiliste au niveau de la population de neurones reste la même, mais le niveau d’activité global (en termes de fréquence de décharge) diminue, en accord avec par exemple les mesures dans le cortex infero-temporal chez le macaque [Assad95]. Nous avons détaillé ces résultats dans une publication commune [KaplanKhoei14]. Il est important de noter que ce type d’implantation est basé sur une connectivité anisotropique qui n’avait jamais —à notre connaissance— été explorée. Ce modèle a fait partie des modèles sélectionnés dans BrainScaleS pour être implantés finalement sur les micro-circuits neuromorphiques à grande échelle.

Un aspect complémentaire à la simulation est l’étude du décodage de l’activité neurale. En collaboration avec Giacomo Benvenutti (thèse BrainScaleS) et Frédéric Chavane (NeOpTo-INT), nous avons étudié des modèles statistiques qui nous permettent d’extraire l’information visuelle de populations de neurones. En premier lieu, une telle approche permet de consolider les bases théoriques qui permettent de caractériser l’activité neurale, tant au niveau de la statistique de fréquence de tir des neurones que pour la paramétrisation des courbes de sélectivité neurale, par exemple en fonction de l’orientation et de la direction. Aussi, une telle caractérisation nous permettent de valider des modèles de décodage et de représentation de l’information dans l’activité neurale (comme par exemple [Jazayeri06]) et ainsi de boucler le lien avec l’implantation neurale de tels processus.

Pour ce faire nous avons recruté une étudiante en post doctorat, Wahiba Taouali, pour mieux comprendre les fondements mathématiques du décodage neural. Une première tâche a consisté à quantifier les différentes sources de bruit dans les enregistrements neurophysiologiques obtenus au laboratoire. En particulier, nous avons développé une méthodologie mathématique permettant de caractériser la variabilité des trains de potentiels d’action. Ce test statistique novateur nous a permis de montrer que la variabilité intrinsèque aux enregistrements s’établissaient de façon différentielle de la rétine, au thalamus et aux aires corticales supérieures. Ces travaux ont fait l’objet d’une publication dans un journal à comité de lecture [Taouali16].

Une fois que ces principes du codage neural étaient mieux compris, nous les avons appliqués à des enregistrements effectués dans le cortex visuel primaire évoqué par une barre en mouvement. Grâce à la méthode de décodage neural nous avons pu valider l’hypothèse selon laquelle dans la représentation neurale, une population de neurones peut coder de façon efficace la position et l’orientation d’une barre en mouvement. En particulier, une signature particulière du codage neural montre qu’une barre en mouvement selon une trajectoire prédictive anticiper le long de son mouvement [Taouali15vss; Taouali16areadne; Perrinet19nccd].

### 3.3.2 Corrélats neurals du mouvement apparent

Appliquons une approche similaire à une autre illusion visuelle : Lorsque deux points stationnaires clignent à des positions et à des moments rapprochés l’un de l’autre, l’observateur peut percevoir une sensation de mouvement. Ce processus transforme la présentation d’un motif discret en un motif continu et est à la base des processus qui nous font percevoir au cinéma une image animée alors que nos yeux ne ressentent qu’une succession rapide d’images statiques. Cette illusion visuelle est appelée mouvement apparent et peut persister sur une distance spatiale relativement longue (supérieure à la taille caractéristique de la RF d’un neurone dans le cortex visuel primaire, V1). Comme dans l’étude ci-dessus pour le FLE, nous pensons que ce mouvement apparent à longue portée (long-range Apparent Motion, lrAM) peut s’expliquer par des processus prédictifs. En raison des caractéristiques dynamiques du lrAM, une implémentation neurale de cette illusion peut consister en la propagation d’informations visuelles par des interactions intra-corticales. En particulier, ces interactions latérales peuvent évoquer des vagues d’activité dans V1 qui peuvent moduler l’intégration de l’information sensorielle provenant des connexions thalamo-corticales. Une perspective intéressante est donc d’enregistrer l’activité neurale lors de la présentation du stimulus lrAM. Ceci permet d’évaluer quantitativement pourquoi la superposition de deux points comme dans le lrAM est “plus” que la somme des deux points isolés.

Dans une étude récente [Chemla19], nous avons utilisé la VSDI pour enregistrer l’activité du cortex visuel primaire (V1) de macaques éveillés. Y a-t-il une différence entre la réponse au point unique et celle aux deux points ? En effet, les enregistrements VSD permettent d’enregistrer l’activité de populations de neurones V1 qui sont approximativement à l’échelle d’une colonne corticale. De plus, la réponse enregistrée est assez rapide pour saisir la dynamique du stimulus lrAM. Les enregistrements montrent que lorsque l’activité évoquée par le second point atteint V1, une onde corticale suppressive se propage vers la vague rétinotopique évoquée par le premier point. Cela a été mis en évidence en comparant statistiquement la réponse du cerveau à la réponse des deux points isolés. En particulier, nous avons constaté que grâce à cette onde suppressive, l’activité du stimulus cérébral était plus précise, ce qui suggère que cette onde suppressive pourrait servir d’étape de traitement prédictive pour être lue dans les aires corticales en amont.

En particulier, nous avons constaté que l’activité que nous avons enregistrée était bien décrite par un modèle de champ moyen utilisant un contrôle de gain dynamique. Qualitativement, ce modèle reproduit la propagation de l’activité sur le cortex. Il est important de noter que ce modèle a permis de montrer que l’activité observée était mieux ajustée lorsque la vitesse des connexions latérales dans le champ moyen était d’environ  $1m/s$ , une vitesse de propagation qui est de l’ordre de celle mesurée pour les connexions intra-corticales dans le cortex visuel primaire (pour une revue, voir [Muller18]). Un modèle plus fonctionnel (probabiliste) a également montré que la vague corticale suppressive permettait de lever l’ambiguïté sur le stimulus en supprimant les alternatives improbables. Par conséquent, (1) les interactions latérales sont essentielles pour générer des vagues d’activité en mouvement à la surface du cortex et (2) ces vagues aident à affiner la représentation du stimulus en entrée. Cela correspond à la mise en œuvre d’un processus prédictif utilisant une connaissance *a priori* des objets visuels en mouvement régulier.

### 3.3.3 Résumé

En résumé, nous avons vu qu’il est possible d’étendre le traitement prédictif aux cartes topographiques. En particulier, les calculs qui en résultent sont particulièrement adaptés à la vision. Nous avons montré (voir 2) un modèle qui représente (à un instant donné) différentes variables (ici “Source” et “Cible”). Dans un modèle plus réaliste, l’activité neurale est plus susceptible de former des représentations intermédiaires entre les représentations passées, présentes et futures [Glaser18] et à différents niveaux d’adaptation comme l’illustre le stimulus lrAM [Chemla19]. En conséquence, de tels processus sont observés phénoménologiquement comme la propagation de l’information neurale tangentielle à la surface corticale, modulant dynamiquement les flux convergeant sur l’aire corticale (“feed-forward” vs “feed-back”). En particulier, la question reste ouverte de savoir si de tels calculs neuraux pourraient être mis en œuvre par des vagues d’activité sur la surface corticale [Muller18].

### 3.4 Productions scientifiques sélectionnées

Je présente ici un choix de 10 productions scientifiques pour les 10 derniers semestres d'activité. Les codes dans la marge correspondent à ceux utilisés dans la liste complète des publications.

**A50 Ladret22**

- Cette publication en révision (deuxième révision) combine 1/ une création de stimuli basée sur un modèle, 2/ des enregistrements neuro-physiologiques basés sur ce protocole novateur et 3/ une méthodologie nouvelle de décodage. Basés sur ces innovations théoriques et expérimentales, nous avons mis en évidence le rôle de la précision dans le traitement dynamique de l'information, et donc le rôle potentiel de processus prédictifs dans des données biologiques.

**A48 Grimaldi22pami**

- Nous avons combiné dans cette publication une approche d'ingénierie des systèmes neuromorphiques et une approche de modélisation des neurosciences computationnelles. L'objectif est de permettre une catégorisation rapide d'objets à partir de flux d'événements. Nous utilisons les données de caméras événementielles et une innovation a été de caractériser de manière extensive l'efficacité de ce réseau en fonction de différents paramètres d'entrée. La nouveauté était de rapprocher ce type de modèle des réseaux événementiels ou des réseaux de neurones à impulsions.

**A47 Chavane22**

- Dans cette revue de l'état de l'art sur l'anatomie du cortex visuel primaire, nous avons proposé une hypothèse novatrice sur l'organisation de la formation sur la surface de cette aire.

**A46 BoutinFranciosiniChavaneRuffierPerrinet20**

- Ce travail étend l'architecture habituellement rencontrée en apprentissage profond en incluant des processus prédictifs. Un résultat principal est d'obtenir après apprentissage un réseau dont la structure est explicable, un résultat majeur par rapport aux réseaux profonds classiques. Un second résultat est de montrer le parallèle entre ce type de réseaux et ce que l'on peut observer dans le cortex visuel primaire.

**A44 Dauce20**

- L'objectif de ce travail est d'introduire des processus actifs dans les modèles actuels de traitement d'images. Inspirés par ce qui peut être observé dans les systèmes biologiques, nous avons introduit des saccades oculaires pour résoudre le problème conjoint de l'identification et de la localisation d'objets dans les images. Le résultat principal est de pouvoir réaliser cette tâche avec beaucoup moins de ressources informatiques.

**A43 PasturelMontagniniPerrinet20**

- Nous avons introduit un nouveau paradigme dans lequel les protocoles expérimentaux peuvent changer à des moments imprévisibles. Ce travail combine une analyse théorique et des résultats comportementaux qui actualisent nos connaissances sur le traitement de la volatilité dans les données comportementales.

**A37 KhoeiMassonPerrinet17**

- Dans cette publication de neurosciences computationnelles, nous avons montré que des principes de propagation dynamique sur la surface du cortex visuel pouvait permettre de donner une solution plus simple à des problèmes classiques de neurosciences de la vision. En particulier nous avons montré qu'un principe prédictif qui accumule de la formation le long d'une trajectoire d'un objet en mouvement permettait de résoudre le problème de l'ouverture.

**A33 PerrinetAdamsFriston14**

- Cette publication permet de donner une un cadre théorique solide sur un problème majeur en neurosciences : le traitement des délais dans la chaîne des processus cognitifs. En utilisant le cadre du principe de minimisation de l'énergie libre, nous avons développé une approche théorique qui permet de compenser pour ces délais.



En particulier nous l'avons appliqué à différents problèmes dans la modélisation des mouvements des yeux, que nous avons étendu dans une autre publication à la modélisation du comportement neuro-typique ou dans le spectre schizophrénique.

**B6 Perrinet20**

- Dans ce chapitre d'ouvrage, je fais une revue de l'état de l'art sur les processus prédictifs et notamment comment ceux-ci peuvent permettre de donner des solutions simples à des problèmes neuroscientifiques.

**B4 CristobalPerrinetKeil15bicv**

- Dans ce livre, nous avons regroupé les contributions de chercheurs en neurosciences et en vision par ordinateur pour donner un cadre général d'une compréhension interdisciplinaire de la perception visuelle.

## 4 Enseignement, formation et diffusion de la culture scientifique

### 4.1 Encadrement de thèse et post-doctorants

Actuellement, j'ai l'occasion d'encadrer trois doctorants en tant qu'encadrant principal :

- Jean-Nicolas Jérémie “Bio-mimetic agile aerial robots flying in real-life conditions” (PhD, bourse AgileNeuroBot (ANR-20-CE23-0021), 10/2021-09/2024)
- Antoine Grimaldi “Ultra-fast vision using Spiking Neural Networks” (PhD, APRO-VIS3D grant (ANR-19-CHR3-0008-03), 2020 / 2023)
- Hugo Ladret “A multiscale cortical model to account for orientation selectivity in natural-like stimulations” (direction, bourse AMU, co-direction avec Christian Casanova, en cotutelle avec l'Université de Montréal, 2019 / 2023)

Précédemment, j'ai eu l'occasion d'encadrer des étudiants en direction de thèse, en post-doctorat ou en co-direction de thèse :

- Alberto Arturo Vergani “Visual computations using Spatio-temporal Diffusion Kernels and Traveling Waves” (Post-Doc, 04/2020 - 09/2021)
- Victor Boutin “Controlling an aerial robot by human semaphore gestures using a bio-inspired neural network” (PhD, bourse AMIDEX, 12/2016-03/2020)
- Angelo Franciosini “Trajectories in natural images and the sensory processing of contours” (PhD, bourse PhD program, 2017 / 2021)
- Kiana Mansour Pour “Predicting and selecting sensory events : inference for smooth eye movements” (PhD, 2015 - 2018)
- Jean-Bernard Damasse “Smooth pursuit eye movements and learning : Role of motion probability and reinforcement contingencies” (PhD, 2014-2017)

### 4.2 Participation à des activités grand public

Je participe ou initie de nombreuses rencontres avec le grand public (cf. <https://laurentperrinet.github.io/project/tout-public/>) et récemment :

- Article de dissemination sur le hasard dans “The conversation” (6200 lectures au 16 Février 2022).
- Participation à une présentation “stand up” des NeuroStories : conférence invitée “Le temps des sens”.
- Article de dissemination sur le temps dans la perception dans “The conversation” (11600 lectures au 16 Février 2022).
- Participation à des activités de dissémination aux des Journées de Neurologie de Langue Française (JNLF) : conférence invitée “Des illusions aux hallucinations visuelles : une porte sur la perception”.
- Article de dissemination sur les illusions visuelles dans “The conversation” (12500 lectures au 16 Février 2022).
- Participation à des activités grand public : Rencontre avec les collégiens marseillais, fête de la science, participation à un jury autour de la société, la science et le cinéma.

### 4.3 Collaboration artistique

En parallèle avec les actions grand public, je développe une collaboration active avec un artiste plasticien, Etienne Rey (friche Belle de Mai, Marseille, voir <https://laurentperrinet.github.io/project/art-science/>). Nous avons produit plusieurs actions, entre autres :

- “Horizon Faille” – interstices, à Caen, 2021
- “Sans gravité – une poétique de l'air” – Ardenome à Avignon, 2019
- “Instabilité (series)” @ Art-O-Rama, Installation avec sérigraphie, dessin mural, lumière, 2018
- projet “TRAMES” présentation à la Fondation Vasarely (Aix), 2016

## 4.4 Enseignement

Cours magistraux de Neurosciences Computationnelles en troisième année de licence Sciences et Humanités et dans le cadre du programme de thèse Marseillais en Neurosciences. 2019  
2018

- An introduction to the field of Computational Neuroscience ,
- Probabilities, the Free-energy principle and Active Inference.

J’ai récemment pris part à une école d’été organisée en janvier 2019 par le centre de neurosciences computationnelles de Valparaiso au Chili. Les thèmes abordés au cours de cette école étaient :

- adaptation comportementale,
- compensation des délais,
- modélisation Bayésienne,
- tutoriel modélisation de réseaux à spikes.

## 5 Transfert technologique, relations industrielles et valorisation

### 5.1 Contrats et collaborations

Au cours de 10 derniers semestres, j’ai eu l’occasion de collaborer sur plusieurs contrats de niveau national (ANR) et international (cf <https://laurentperrinet.github.io/#grants>) :

Actuellement, je suis impliqué dans les contrats suivants

- soit à titre d’investigateur principal :
  - ANR AgileNeuroBot (Co-ordinateur principal) : “Robots aériens agiles biomimétiques pour le vol en conditions réelles” (2021/2024)
  - APROVIS3D : “Aprovis3D : Event-Based Artificial Intelligence” (2019–2023, coordination globale du projet par Jean Martinet, université de Nice).
  - ANR PRIOSENS (2021/2024) : “Modelling behavioural and neuronal data within the active inference framework” avec Anna Montagnini,
- soit à titre de collaborateur :
  - ANR ShootingStar (2021/2024) avec Frédéric Chavane,
  - ANR ACES (2021/2024) : “CAssignment of credit and constraints on eye movement learning” avec Anna Montagnini,
  - ANR RubinVase (2021/2024) avec Dario Prandi et Luca Calatroni.

Précédemment, j’ai eu l’occasion de collaborer sur les contrats suivants durant cette période :

- PhD DOC2AMU : An Excellence Fellowship, H2020 (Excellence Scientifique) : Actions Marie Skłodowska-Curie (IF, ITN, RISE, COFUND) (2016–2019)
- PhD ICN A grant from the Ph.D. program in Integrative and Clinical Neuroscience (PhD position, 2017 / 2021).
- SpikeAI : lauréat du Défi Biomimétisme (2019) “Algorithmes événementiels d’Intelligence Artificielle / Event-Based Artificial Intelligence” (2019).
- ANR Horizontal-V1 (2017–2021) : “Connectivité Horizontale et Prédiction de Cohérences dans l’Intégration de Contour” avec Yves Fregnac et Frédéric Chavane,

### 5.2 Développements de logiciels

Nous développons plusieurs lignes de recherche pour appliquer nos résultats à des problèmes concrets, sous forme de logiciels open source :

#### 5.2.1 Mouvements des yeux et mouvement

- AnEMo : traitement du signal pour l’analyse des mouvements des yeux [Pasturel18anemo],
- MotionClouds : génération de textures pour la perception du mouvement [Sanz12 ; Vacher15nips ; Vacher16],

- LeCheapEyeTracker – <https://github.com/laurentperrinet/CatchTheEye> : Oculomètre minimal utilisant l'apprentissage profond ;

### 5.2.2 Biologically-Inspired Computer Vision

- openRetina : caméra événementielle minimale,
- SparseHebbianLearning : apprentissage non-supervisé d'images naturelles [Perrinet10shl ; Perrinet19hulk],
- Simple Library for Image Processing : techniques de traitement de l'image, utilisé notamment dans [Perrinet15bicv ; Ravello16droplets ; PerrinetBednar15 ; Perrinet15eusipco ; Perrinet16EUVIP],
- LogGabor : représentations multi-échelles des contours [Fischer07 ; Fischer07cv],
- SparseEdges : codage épars (parcimonieux) d'images naturelles [Perrinet15bicv ; PerrinetBednar15],
- MotionParticles : prédiction dynamique par filtrage particulaire (permet de reproduire [Perrinet12pred ; Khoei13jpp ; KhoeiMassonPerrinet17]).

### 5.2.3 Promotion du logiciel libre

- Je participe à différentes initiatives afin de promouvoir les pratiques du logiciel libre
- écriture régulière d'un blog scientifique,
  - participation à des réseaux sociaux à des fins de dissémination comme twitter, stackOverflow, instagram ou gitHub.

## 5.3 Expertise scientifique

J'ai participé au développement de différentes entreprise de type "start-up" dans le cadre d'une autorisation de cumul :

- en 2019-2020 : missions d'expertise scientifique avec Arnaud Malvache à Unistellar, Marseille.
- 2019-2020 : missions d'expertise scientifique en collaboration avec Sid Kouider à NextMind, Paris.

## 6 Encadrement, animation et management de la recherche

Depuis Janvier 2022, je suis nommé à la commission interdisciplinaire (CID) 51 : Modélisation mathématique, informatique et physique pour les sciences du vivant.

Outre ces responsabilités scientifiques, je participe à l'animation scientifique sous d'autres formes. Tout d'abord pour l'évaluation de la recherche par les chercheurs en tant que membre d'un comité éditorial ou en temps que relecteur. Je développe aussi des collaborations internationales et en même temps dans la vie sociale de l'organisme :

- Scientific reports (Nature group) Membre du comité éditorial
- Relecteur dans de nombreuses revues et conférences, voir <https://publons.com/author/1206845/laurent-u-perrinet#profile>
- Membre de différents jurys internationaux (BE, NL) pour des évaluations de projets,
- Membre élu CLAS GLM de Marseille-Joseph Aiguier/Timone, responsable de la petite enfance.

## 7 Objectifs / Projet de recherche de janvier 2022 à janvier 2027

Au sein du système nerveux central, les aires visuelles sont essentielles pour transformer le signal lumineux brut en une représentation qui transmet efficacement des informations sur l’environnement. Ce processus est contraint par la nécessité d’être robuste et rapide. En effet, il existe d’un côté une grande variété de changements possibles dans les caractéristiques géométriques de la scène visuelle et de l’autre, il existe une urgence continue d’être capable de répondre à chaque instant et le plus rapidement possible au flux sensoriel entrant, par exemple pour entraîner un mouvement des yeux vers l’emplacement d’un danger potentiel. Des décennies d’études en neurophysiologie et en psychophysique aux différents niveaux de vision ont montré que de nombreuses facettes de ce système tirent profit des connaissances *a priori* sur la structure de l’information visuelle, telles que la régularité dans la forme et le mouvement des objets visuels. Ainsi, le cadre du *traitement prédictif* offre une théorie unifiée pour expliquer une large variété de mécanismes visuels. Toutefois, **il nous manque encore une approche normative globale unifiant les processus visuels.**

**Le projet de recherche que je développe explore la dynamique du traitement prédictif dans le système visuel, de la rétine à l’action.** Afin de le définir, nous passerons rapidement en revue ici quelques approches récentes et prometteuses auxquelles j’ai eu la chance de contribuer. Tout d’abord, nous décrirons l’inférence active, une forme de traitement prédictif doté de la capacité d’échantillonner activement l’espace visuel. Ensuite, nous étendrons ce paradigme au cas où l’information est distribuée sur une topographie, comme c’est le cas pour les aires visuelles organisées rétinotopiquement. En particulier, nous comparerons ces modèles à la lumière de données neurophysiologiques récentes montrant le rôle des vagues d’activation dans la formation du traitement visuel. Enfin, **je vais proposer un programme de recherche original pour comprendre comment ces modèles fonctionnels peuvent être mis en œuvre au niveau neural sous la forme de cartes topographiques adaptatives de micro-circuits prototypiques.** Ceux-ci permettent de séparer les différents flux d’information, depuis l’erreur de prédiction sensorielle à l’erreur d’anticipation de la rétroaction. Néanmoins, la conception d’un tel circuit de traitement prédictif générique est un champ de recherche ouvert et nous énumérerons quelques implémentations possibles à l’aide de réseaux de neurones biomimétiques.

### 7.1 Motivation : Dynamique des calculs neuronaux sous-jacents au traitement visuel

La vision, c’est-à-dire la capacité de donner un sens à l’environnement lumineux, est traditionnellement considérée comme une séquence d’étapes de traitement allant de l’entrée rétinienne à une représentation de niveau supérieur, permettant éventuellement une action. On pense souvent que cette séquence d’étapes de traitement, ou “pipeline”, est mise en œuvre par un processus “feed-forward” (en avant) dans les voies visuelles, à travers le thalamus et ensuite dans les aires visuelles du cortex cérébral. Un tel modèle de vision est suffisant pour expliquer la simple détection du caractère imprimé que vous êtes en train de regarder, et donc pour la lecture d’une phrase complète. En effet, une telle capacité implique des processus de bas niveau rapides et inconscients. Il est important de noter que cette aptitude chez l’homme est également largement robuste aux changements de luminance (comme une ombre sur cette page) ou aux déformations géométriques, comme lors de la lecture de ce texte dans une perspective inclinée. Plus généralement, la vision biologique complètera correctement l’image d’un mot avec des lettres manquantes ou avec des détections ambiguës ou incorrectes dues à une occlusion ou un chevauchement. Une telle robustesse est caractéristique des systèmes biologiques, d’où son utilisation comme test de Turing pour les algorithmes de sécurité tels que les CAPTCHAs. De manière générale, les modèles de vision tels qu’ils sont mis en œuvre dans les ordinateurs peuvent apprendre de telles tâches sur des ensembles de données très précis. Par exemple, des réseaux de neurones artificiels de type “apprentissage profond” ont récemment prouvé qu’ils pouvaient dépasser la performance humaine sur une tâche de catégorisation d’images statiques [NIPS2012\_4824]. Toutefois, ces mêmes processus

sont facilement surpassés par un enfant de 6 ans lorsqu'il s'agit d'un contexte écologique, flexible et générique. En allant encore plus loin, la vision humaine se caractérise aussi par des processus de plus haut niveau et permet des prédictions prospectives telles que celles révélées par l'imagerie mentale - et constitue la pierre angulaire de la créativité, ou de son *imagination*. La vision est donc un processus extrêmement complexe, et les étapes de traitement ne sont sûrement pas indépendantes. En fait, le plus surprenant au sujet de la vision est la facilité avec laquelle les personnes voyantes peuvent apprendre à exercer ces capacités. Pour reformuler **Wigner90empty citation**, "l'efficacité déraisonnable de la vision dans le monde naturel" nous invite à nous concentrer sur cette capacité cognitive pour une meilleure compréhension du cerveau en général.

Anatomiquement, la vision est le résultat de l'interaction de réseaux neuronaux qui sont organisés en une hiérarchie d'aires visuelles. Chaque aire visuelle est un processus dynamique, depuis sa première étape, la rétine, jusqu'aux aires visuelles efférentes qui contribuent à former une représentation parallèle et distribuée du monde visuel. De plus, cette organisation est largement auto-organisée et très efficace sur le plan métabolique. Pour comprendre le fonctionnement d'un réseau aussi complexe d'aires visuelles, il a été proposé que ce système soit organisé de manière à *prédire* efficacement les données sensorielles [Attneave54]. Cette approche écologique [Atick92] permet d'expliquer de nombreux aspects de la vision comme traitement prédictif. Une telle approche prend différentes formes telles que la réduction de la redondance [Barlow61], la maximisation du transfert d'information [Linsker90] ou la minimisation de l'énergie métabolique. La formalisation de ces stratégies d'optimisation en langage probabiliste peut être englobée dans le cadre du "cerveau Bayésien" [Knill04]. Plus généralement, il est possible de relier ces différentes théories dans un cadre unique, le principe de l'énergie libre (Free-Energy Principle, FEP) [Friston10]. Ce principe constitue un changement de paradigme crucial pour l'étude des processus prédictifs tant au niveau philosophique que scientifique. La clé de ce principe est la notion que, connaissant les processus qui ont généré l'image visuelle et le modèle de génération interne qui permet sa représentation, les processus prédictifs profiteront de la connaissance *a priori* pour former une représentation optimale de la scène visuelle [Rao99]. Cette connaissance constitue une représentation explicite (probabiliste) de la structure du monde. Par exemple, une image composée de bords sera comprise à un niveau supérieur dans la hiérarchie visuelle en utilisant la connaissance *a priori* du lien entre les bords individuels pour former une représentation des *contours* des objets visuels. Dans le domaine temporel, la connaissance des transformations géométriques telles que le mouvement des objets visuels aidera à prédire leur position future et à suivre les différents segments du mouvement, mais aussi à représenter les contours invariablement à ce mouvement.

Cependant, il y a des limites et des contraintes à l'efficacité de la vision. Premièrement, l'information lumineuse peut être bruitée et ambiguë, par exemple dans des conditions de faible luminosité. Cela contraint le système à être robuste face aux incertitudes. Cela met en évidence un avantage clé du traitement prédictif, car il implique l'apprentissage d'un modèle génératif des données sensorielles. D'une part, en représentant explicitement la précision des variables (l'inverse de la variance inférée de sa valeur), on peut intégrer de manière optimale des informations distribuées, même dans le cas où cette incertitude n'est pas uniforme et évolue dynamiquement dans le système. D'autre part, un modèle génératif permet de représenter explicitement les transformations des données (comme une transformation géométrique de l'image comme une translation ou une rotation) et donc de faire des prédictions sur les états futurs. Deuxièmement, les réseaux neuronaux ont des capacités limitées de transfert de l'information et ont toujours besoin d'un certain délai pour transmettre et traiter l'information. Chez l'humain, par exemple, le délai de transmission de l'information rétinienne au cortex est d'environ 50 ms, tandis que la latence minimale pour effectuer une action oculomotrice est d'environ 50 ms [Kirchner06] (voir [Lamme00] pour des valeurs équivalentes chez le singe). Bien que cela limite naturellement la capacité du système visuel, nous profiterons ici de ces délais pour disséquer les différents processus visuels. **En particulier, nous nous concentrerons dans ce programme de recherche sur ma contribution dans l'étude du rôle de ces contraintes temporelles sur la dynamique des processus prédictifs.**

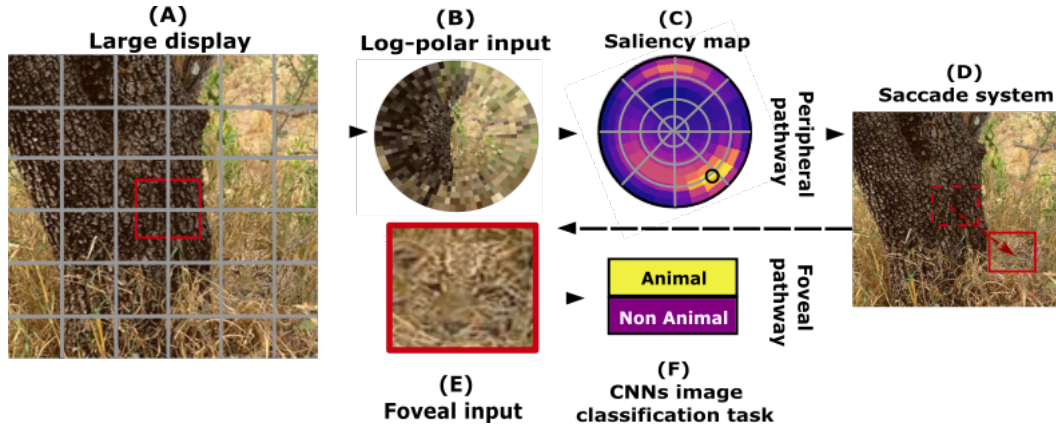


FIGURE 3 – Dans un modèle à double voie du système visuel [Dauce20], la voie périphérique (rangée supérieure) est appliquée à une image de grande taille provenant d’une scène naturelle (A) : Il est d’abord transformé en une entrée log-polaire rétinotopique (B), puis on apprend à renvoyer une “carte de sillance” (C). Cette dernière déduit, pour différentes positions dans la cible, la valeur de précision prédite qui peut être atteinte par la voie fovéale, imitant la voie “Où” utilisée pour la localisation globale. La position présentant la meilleure précision alimentera un système de saccades (D), ajustant le point de fixation à l’entrée de la voie fovéale (ligne du bas). Il prend un sous-échantillon (E) de la large image (A), sur laquelle une catégorisation est effectuée (F), imitant la voie “Quoi”. La modélisation de cette architecture à double voie permet d’offrir un modèle efficace de la recherche visuelle en tant que vision active. En particulier, elle permet de combler les lacunes des CNN en matière de performances physiologiques. Dans le futur, nous espérons appliquer ce modèle pour mieux comprendre les pathologies visuelles dans lesquelles il existerait une déficience de l’une des deux voies.

Pour illustrer le défi de représenter un signal dynamique, prenons l’exemple de l’enregistrement d’un ensemble de cellules neurales dans certaines aires visuelles. Supposons que ces enregistrements soient évoqués par un signal visuel analogique (comme un signal lumineux projeté sur une population de cellules sensorielles rétinienne) et que l’on puisse extraire les temps exacts des événements de décharge pour une population de cellules. Nous pouvons ensuite choisir d’afficher ces données sur une figure sous forme de “tracé temporel”, c’est-à-dire de montrer le moment des spikes pour chacune des cellules identifiées. Le temps est donc relatif à celui de l’expérimentateur et est donné grâce à une horloge externe : Il est montré *a posteriori*, c’est-à-dire après l’enregistrement. En général, cette définition d’un temps absolu a d’abord été formalisée par Newton et définit la plupart des lois de la physique, utilisant le temps comme un paramètre externe. Mais il n’y a encore aucune preuve que les neurones auraient accès à une horloge centrale qui donne accès à l’heure physique absolue. Au contraire, les réponses neurales sont uniquement contrôlées par la distribution *actuelle* des gradients électrochimiques sur leur membrane, potentiellement modulés par les cellules voisines. Une telle notion du temps est propre à chaque neurone et à son environnement. En conséquence, la dynamique du réseau est largement asynchrone, c’est-à-dire que le “chronométrage” est décentralisé. De plus, cette notion locale de temps (de traitement) est *a priori* disjointe du temps externe qui est utilisé pour représenter le signal visuel. Une telle observation est essentielle pour comprendre les principes qui guident l’organisation des processus visuels : Une théorie neurale des processus prédictifs ne peut être définie que dans ce temps local, en utilisant uniquement les informations disponibles localement et à l’instant présent. En particulier, nous proposerons que les processus neuronaux en vision visent à “prédire le présent” [Changizi08] en utilisant un modèle générateur interne de la scène visuelle et en utilisant des données sensorielles pour valider cette représentation interne.

Ce programme de recherche passe en revue ces approches de traitement prédictif dynamique pour la vision à différentes échelles d’analyse, de l’ensemble du système aux représentations intermédiaires et enfin aux neurones (en suivant dans un ordre décroissant

les niveaux d'analyse de **Marr83empty citation**). Tout d'abord, nous appliquerons la FEP à la vision en tant qu'approche normative. De plus, les représentations visuelles doivent prendre en charge les transformations géométriques (comme le mouvement d'un objet visuel) mais aussi les modifications sensorielles, comme les mouvements des yeux. En étendant le principe précédent à la capacité d'échantillonner activement les données sensorielles, nous définirons l'inférence active (IA) et illustrerons son rôle potentiel dans la compréhension de la vision, ainsi que des comportements comme les mouvements oculaires (voir Figure 3 pour une implémentation que nous proposons). Ensuite, nous l'étendrons pour comprendre comment de tels processus peuvent être mis en œuvre sur des cartes rétinotopiques. Ces données seront ensuite comparées aux données neurophysiologiques et nous examinerons les implémentations possibles de tels modèles dans les réseaux de neurones impulsifs. En particulier, nous passerons en revue quelques modèles de micro-circuits élémentaires et détaillerons quelques règles potentielles pour apprendre la structure de leurs connexions d'une manière non supervisée. Nous pourrions ainsi conclure en dessinant les perspectives de résultats de ce programme de recherche.

## 7.2 Problèmes ouverts des processus visuels prédictifs

Dans [PerrinetAdamsFriston14], nous avons étudié la dynamique du traitement prédictif à l'échelle macroscopique, c'est-à-dire en considérant les aires (corticales) comme des noeuds d'un graphe de dépendance. Dans la section 3.2.4, nous avons étendu ces modèles à la topographie de chaque aire visuelle. À une échelle encore plus fine que cette échelle mésoscopique intermédiaire se trouve l'échelle microscopique des cellules neurales. Pour mieux comprendre les mécanismes du traitement prédictif, je propose d'affiner maintenant la granularité de la modélisation à l'échelle neurale pour définir une hiérarchie de cartes de micro-circuits prototypiques. En particulier, en plus de la nature asynchrone de la représentation neurale que nous avons explorée ci-dessus, la communication entre neurones a la propriété d'être basée sur des événements. En effet, la grande majorité des cellules neurales à travers le monde vivant communiquent en utilisant des impulsions courtes et prototypiques appelées potentiels d'action ou *spikes*. Dans cette section, nous proposons trois problèmes ouverts qui sont soulevés lors de la modélisation de tels réseaux de neurones impulsifs (Spiking Neural Networks, SNN), en particulier dans le contexte du traitement prédictif.

### 7.2.1 Les défis de la représentation de l'information visuelle dans les réseaux de neurones impulsif (SNN)

Après les premières générations d'ANN, les algorithmes actuels d'apprentissage machine tels que les algorithmes d'apprentissage profond (Deep Learning, DL) constituent une percée qui a formé une deuxième génération d'ANNs. Les SNN constituent une troisième génération potentielle [Ghosh09]. En effet, la représentation événementielle présente de nombreux avantages par rapport aux ANNs classiques. Par exemple, au lieu de répéter tous les calculs pour chaque couche, canal et pixel d'un ANN hiérarchique, et pour lesquels des GPU gourmands en énergie sont nécessaires, il suffit dans un SNN d'effectuer les calculs seulement sur les événements émis par des unités actives (au moment du spike). En particulier, un domaine de recherche en plein essor consiste à développer des matériels dédiés, comme les puces neuromorphiques, qui permettraient d'étendre le volume effectif des calculs au-delà des dernières générations de semi-conducteurs classiques (CPU, GPU) qui atteignent les limites de la loi de Moore.

La nature discrète de l'adressage des neurones, d'une part, et la nature analogique de la synchronisation des spikes, d'autre part, sont essentielles dans ce nouveau type de représentation. Des résultats notables utilisant de telles architectures ont été obtenus dans la classification en temps réel et la fusion de capteurs [Oconnor13] et dans la reconnaissance de formes [Lagorce17; Grimaldi21cbmi; Grimaldi22pami]. En effet, une propriété importante des SNNs est la capacité d'encoder dynamiquement une variable interne latente (le potentiel de membrane en neuro-physiologie) et d'émettre un spike quand (et seulement quand) un seuil défini en interne est atteint. Ceci définit chaque neurone spikant comme



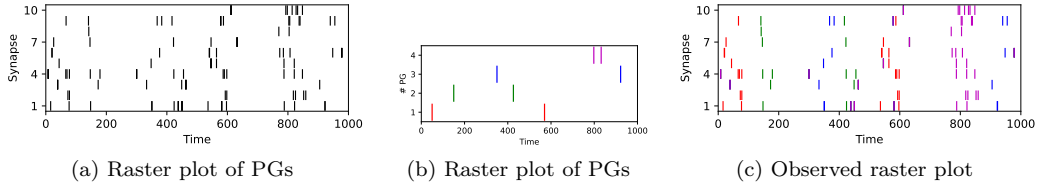


FIGURE 4 – Un exemple de “raster plot” synthétique est généré dans (a) comme la combinaison de PGs dont le timing et l’identité sont déterminés dans (b), ici une superposition de 4 PGs projetés sur l’espace synaptique de 10 neurones et à une fréquence moyenne de 3 Hz pendant 1000 ms. En observant les spikes de (a), le modèle décode l’identité et le timing des PGs dans (b), ce qui nous permet de colorer les spikes dans (c).

un intégrateur (similaire aux neurones classiques), mais aussi potentiellement comme un détecteur de synchronisation [Perrinet02]. Cette capacité à moduler le traitement en fonction de la synchronisation relative des spikes présynaptiques constitue un nouveau paradigme pour les calculs neuraux [Paugam12]. Cela montre en particulier que l’équilibre du flux des spikes excitateurs et inhibiteurs entrants est crucial pour maximiser l’efficacité de ces SNN [Hansel12]. Nous envisageons à court terme d’utiliser des méthodes d’analyse de “raster plots” pour mieux élucider le rôle des événements dans le code neural (voir Figure 4 pour une telle hypothèse).

### 7.2.2 Le rôle des vagues d’activité corticales dans le traitement dynamique de l’information visuelle

Un autre point crucial dans le décryptage des mécanismes de traitement prédictif est donné par l’anatomie fonctionnelle. En effet, dans le cortex visuel primaire (V1) comme dans les autres aires corticales, le réseau neural est très majoritairement récurrent avec un nombre médian de 10000 connexions par neurone. Étonnamment, 95% de ces connexions se produisent dans un rayon de 2mm (singé macaque) [Markov13]. Cela suggère qu’une majorité de ressources neurales est consacrée aux communications intra-régionales. Un rôle fonctionnel supposé de ce réseau dense est de générer des vagues d’activité qui modulent la force et la dynamique de l’activité neurale “feed-forward” [Muller18]. Nous avons vu son rôle potentiel dans la désambiguïsation du mouvement [Chemla19] et il a également été démontré qu’il facilite l’accumulation progressive de l’information visuelle [Bringuier99]. Auparavant, nous avons modélisé avec succès un tel processus prédictif [Perrinet12pred ; Khoei13jpp ; KhoeiMassonPerrinet17], et l’avons mis en œuvre dans un SNN [Kaplan13].

Un “Saint Graal” dans cette direction est de trouver des micro-circuits canoniques pour le codage prédictif [Bastos12]. Cela découle de l’observation qu’à travers les espèces et les aires, le cortex semble suivre une structure prototypique en couches. Dans le cas particulier de V1, alors que l’entrée thalamique atteint principalement la couche granulaire (intermédiaire), un flux de feed-forward se propage principalement aux couches efférentes à travers les couches supra-granulaires tandis que le feed-back est en majorité conduit par les couches infra-granulaires. Cette ségrégation anatomique pourrait correspondre à différents types de signaux dans le codage prédictif, respectivement des états attendus et des erreurs de prédiction [Bastos12]. De tels micro-circuits de base ont été utilisés pour expliquer la réponse des neurones V1 aux scènes naturelles [Kremkow16] en utilisant un mécanisme push-pull. Des réseaux fonctionnels utilisant un apprentissage dépendant du temps des spikes ont aussi été utilisés pour expliquer des réponses obtenues chez le chat [Ladret19sfn]. Il n’en demeure pas moins qu’il est difficile de savoir comment un tel circuit peut émerger de principes computationnels généraux.

### 7.2.3 Propriétés intégratives des aires corticales

Une autre perspective intéressante est la nature intégrative des calculs neuraux. Bien que l’on croit souvent que les neurones représentent la combinaison de caractéristiques

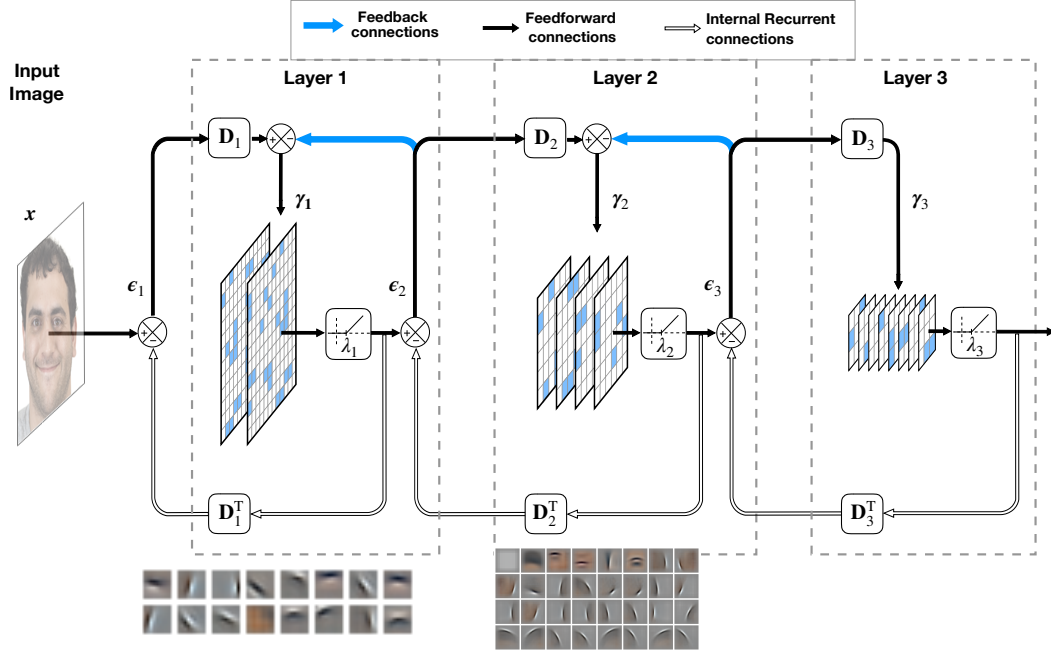


FIGURE 5 – Dans [BoutinFranciosiniChavaneRuffierPerrinet20], nous proposons un modèle de traitement hiérarchique de l'information visuelle (ici de 3 couches). De façon similaire à des réseaux de neurones convolutionnels classiques, chaque couche est constituée de canaux auxquels sont attribués des champs récepteurs. Toutefois, contrairement aux réseaux classiques, l'apprentissage est ici bio-mimétique, c'est-à-dire qu'au lieu d'opérer une rétro-propagation (globale) du gradient, les opérations sont toutes locales et s'opèrent de proche en proche. Cet apprentissage non-supervisé est possible par l'introduction d'une contrainte de régularisation assurant la parcimonie de la représentation. On observe l'émergence de champs récepteurs sensibles à des bords orientés dans la première couche, de façon similaire à ce qui est observé dans le cortex visuel primaire. De façon plus surprenante, quand on apprend le réseau sur une base d'images de visages, on observe l'émergence de cellules sensibles à des composantes indépendantes : bouche, yeux, contours allongés, ...

visuelles, ce n’est en général pas correct [Tring18]. Au lieu de cela, il a été constaté que l’activité peut devenir plus précise à mesure que les caractéristiques visuelles s’accumulent. Par exemple, [Baudot13] a montré que les neurones de l’aire 17 du chat réagissent plus sélectivement lorsqu’ils se voient présenter des images naturelles (qui consistent localement en une somme de bords) qu’à un seul bord isolé. Récemment, nous avons montré qu’un résultat similaire peut se produire chez les rongeurs dès la rétine [Ravello19]. Sur le plan comportemental, cela correspond également à l’observation chez l’homme que des textures plus complexes entraînent des mouvements oculaires plus robustes [Simoncini12]. Ces phénomènes sont conformes au principe du traitement prédictif selon lequel, en accumulant des informations cohérentes, la probabilité *a posteriori* (et donc la réponse du système) devient plus précise.

Fait frappant, cela se traduit dans l’activité neurale par le fait que, pour un ensemble plus cohérent d’entrées sensorielles, l’activité neurale de la population est plus “parcimonieux” (en anglais, “sparse”) [Vinje02; Baudot13]. Cela s’expliquait déjà par le modèle de codage prédictif de [Rao99] et mis en œuvre dans [Kremkow16] par exemple. Il est important de noter que le principe de codage parcimonieux est lui-même suffisant pour (1) expliquer de façon raisonnée une grande partie des mécanismes de contrôle du gain [Heeger17] et (2) guider l’apprentissage de la connectivité dans une population de neurones, comme dans V1 [Olshausen97; Perrinet10shl; Perrinet15bicv; Perrinet19hulk]. Cela aide à résoudre un problème important, à savoir que le système est auto-organisé et que l’apprentissage de la connectivité ne doit pas être supervisé. Ainsi, les règles de plasticité qui devraient être élaborées dans les SNN devraient utiliser des principes directeurs similaires.

Cependant, il nous manque encore des modèles réalistes d’un tel traitement prédictif visuel. Nous avons construit un modèle simplifié capable de traiter des images statiques [BoutinFranciosiChavaneRuffierPerrinet20]. Il s’agit d’un réseau neural convolutionnel multicouche, où chaque couche comprend à la fois un mécanisme intracortical récursif pour générer des représentations parcimonieuses et la possibilité pour chaque couche d’intégrer des informations (feedback) provenant d’une couche de niveau supérieur, voir Figure 5. La principale nouveauté de ce réseau est qu’il permet l’apprentissage non supervisé des noyaux convolutionnels à chaque couche. Comparés aux réseaux de neurones convolutionnels classiques tels que ceux que l’on trouve couramment dans les architectures d’apprentissage profond, nous avons constaté que les noyaux émergents étaient plus interprétables : Par exemple, en apprenant sur une classe d’images de visages humains, nous avons observé dans la deuxième couche différents neurones sensibles aux caractéristiques du visage comme les yeux, la bouche ou le nez. C’est similaire à ce que l’on trouve dans l’aire corticale du lobule fusiforme, mais d’autres simulations sont nécessaires pour valider l’émergence de cette représentation. Nous avons aussi observé que les connections en retour, en “feedback” avaient une importance significative sur cette émergence [BoutinFranciosiRuffierPerrinet20feedback]. Toutefois, ces simulations sont intensives en calcul et interdisent leur extension à des flux dynamiques sur des architectures informatiques conventionnelles. Une traduction de cet algorithme en un réseau neural impulsionnel serait donc très bénéfique et permettrait de l’appliquer à un flux dynamique d’images.

### 7.3 Résumé et conclusions

En résumé, nous avons examiné dans ce programme de recherche différents modèles de codage prédictif appliqués à la vision. Nous avons vu à l’échelle macroscopique le rôle de la dynamique à l’aide de l’inférence active. En étendant ce modèle à une carte rétinotopique, nous pourrions décrire le rôle fonctionnelle des vagues d’activité pour améliorer la sélectivité à des stimuli visuels (voir section 3.2.4). Cependant, nous avons également montré une limite de ces modèles à l’échelle microscopique (voir section 7.2). En particulier, on ne comprend pas encore, au niveau de la cellule unique, comment (1) l’information est représentée dans l’activité impulsionnelle, (2) quel est le rôle fonctionnel des vagues d’activité sur les surfaces corticales, (3) si un principe d’efficacité commun (comme un codage parcimonieux) pouvait

être utilisé pour guider l'organisation de ces réseaux hautement récurrents dans un circuit universel unique.

Pour approfondir nos connaissances du traitement prédictif en vision (voir section 7.2), il semble donc nécessaire de pouvoir mettre en œuvre des SNN à grande échelle mettant en œuvre des processus visuels complexes. Cependant, les trois différentes échelles anatomiques que nous avons mises en évidence ci-dessus (feed-forward, latéral, feedback) semblent être étroitement couplées et peuvent être difficiles à modéliser séparément. Plus généralement, c'est également vrai pour les échelles que nous avons définies, depuis le macroscopique, jusqu'au mésoscopique et au microscopique. Il est donc très difficile de produire des modèles assez simples pour nous aider à comprendre le traitement sous-jacent [Varoquaux19; Brette19]. Par exemple, après les avoir déduits des principes d'optimisation, tous les modèles que nous avons présentés ici sont préconnectés : Les hyper-paramètres contrôlant l'interconnexion des neurones sont fixes. Bien que nous ayons fourni des simulations montrant le rôle de ces hyperparamètres, il semble nécessaire de mieux comprendre leurs effets relatifs.

En effet, une théorie normative du traitement prédictif ne devrait pas seulement fournir une solution possible (un modèle donné avec un ensemble d'hyperparamètres) mais aussi explorer *toutes les solutions possibles*. Une première méthodologie consiste à avoir une compréhension complète de l'ensemble des modèles à l'aide de l'analyse mathématique. Cependant, cela devient impossible pour des systèmes aussi complexes et l'utilisation d'hypothèses simplificatrices conduit souvent à une complexité superficielle. Un autre moyen consiste à élaborer des stratégies d'adaptation pour explorer l'espace fonctionnel de différents modèles. Ceci peut par exemple être développé à l'aide de techniques d'apprentissage machine telles que la descente stochastique de gradient couramment utilisée dans l'apprentissage profond. Une autre solution prometteuse consiste à explorer des stratégies d'adaptation inspirées de la biologie. Celles-ci existent à différentes échelles temporelles, allant des mécanismes d'adaptation rapide, à un apprentissage plus lent des connexions, ou à l'évolution à long terme des hyper-paramètres. En particulier, on ne comprend pas encore tout à fait comment implanter dans les SNNs une plasticité dépendante du temps de spike. Cela pose un défi futur dans notre compréhension de la science des processus prédictifs en vision qui définira mon prochain programme de recherche.