Modeling visual information processing during locomotion using bio-inspired spiking neural networks

Paul Fricker

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Paul FRICKER

Le 12 décembre 2022

Modélisation du traitement visuel de l'information lors de la locomotion à l'aide de réseaux de neurones impulsionnels bio-inspirés

Ecole doctorale : **CLESCO - Comportement, Langage, Education, Socialisation, Cognition**

Spécialité : **Neurosciences**

Unité de recherche :
CERCO - Centre de Recherche Cerveau et Cognition

Thèse dirigée par
Benoit COTTEREAU et Christophe HURTER

Jury

Mme Samia BOUCHAFA-BRUNEAU, Rapporteure
M. Martial MERMILLOD, Rapporteur
M. Ioan Marius BILASCO, Examinateur
M. Daniel DELAHAYE, Examinateur
M. Benoit COTTEREAU, Co-directeur de thèse
M. Christophe HURTER, Co-directeur de thèse

Résumé en français

Titre : Modélisation du traitement visuel de l'information lors de la locomotion à l'aide de réseaux de neurones impulsionnels bio-inspirés

Mots clés : Réseau de neurones impulsionnels, locomotion, flux optique, apprentissage non supervisé

Le domaine de l'intelligence artificielle est en plein essor depuis de nombreuses années et définit la base de tout apprentissage s'effectuant par des systèmes informatiques. Parmi les méthodes d'apprentissage utilisées se retrouvent les réseaux de neurones artificiels. Le développement de ces derniers était à l'origine grandement motivé par le désir de comprendre les mécanismes biologiques sous-tendant les traitements effectués par le cerveau humain. Au cours des dernières années, les réseaux de neurones se sont de plus en plus complexifiés avec un nombre de paramètres et une consommation énergétique toujours plus élevés. Si les réseaux de neurones profonds ou convolutionnels actuels ont d'excellentes performances pour de nombreuses tâches cognitives, il est désormais difficile de les utiliser comme outil pour comprendre le traitement de l'information chez le vivant. Certains types de réseaux de neurones artificiels restent néanmoins plus ancrés dans la biologie, il s'agit des réseaux de neurones impulsionnels (Spiking Neural Networks ou SNN en anglais). Ils reposent sur des mécanismes de communication observés entre les neurones. Ces derniers se transmettent des informations sous la forme d'impulsions électriques dont la consommation énergétique est très réduite. L'apprentissage au sein de ces SNN peut s'effectuer de façon non-supervisée à partir de règles de plasticité observées chez le vivant comme la Spike-Timing-Dependent Plasticity ou STDP.

Cette thèse a pour objectif de modéliser le traitement visuel lors de la locomotion à partir de SNNs régulés par un apprentissage de type STDP, recevant des données captées par des caméras asynchrones dont le fonctionnement s'inspire directement de celui de la rétine. Une première étude ([Fricker et al., 2022]) se penche sur la création d'un tel réseau pour l'apprentissage des composantes de flux optique. En interfaçant ce SNN avec des entrées de données événementielles d'abord simulées, nous avons pu mettre en évidence les performances du réseau pour l'estimation des différentes composantes du flux optique à travers différentes conditions de bruit et de vitesse. Le même réseau a par la suite été entrainé à partir de données de navigation réelles capturées à l'aide d'une caméra événementielle. Si la tâche s'avère plus difficile malgré un décodage du flux optique convaincant, elle met en évidence le besoin d'un jeu de données à mi-chemin entre des données événementielles simulées et capturées en condition réelle afin de se rapprocher de tâches de navigation naturelle tout en permettant d'en contrôler les données et paramètres.

Une seconde étude s'intéresse à la création d'un tel jeu de données pour l'apprentissage du réseau. Ce jeu de données est obtenu à partir d'un logiciel dédié permettant d'extraire puis de modéliser la structure 3D de différents environnements. Ce logiciel peut ensuite simuler les spikes générés par une caméra virtuelle se déplaçant selon différentes trajectoires au sein de ces environnements. Cette approche permet de mieux contrôler les données reçus par le réseau et d'améliorer ainsi l'apprentissage. Les premiers résultats montrent que les neurones du réseau après apprentissage deviennent sélectifs aux différentes composantes du flux optique tout en restant invariant aux propriétés spatiales. La méthode proposée constitue donc une base solide pour le traitement du flux optique à partir de réseau de neurones impulsionnels et au sein de différents contextes environnementaux.

Résumé en anglais

Title: Modeling visual information processing during locomotion using bio-inspired spiking neural networks

Keywords: Spiking neural networks, locomotion, optic flow, unsupervised learning

Artificial intelligence has been booming for many years and defines the basis for all learning carried out by computer systems. Among the learning methods used are artificial neural networks. The latter's development was initially greatly motivated by the desire to understand the biological mechanisms underlying the treatments carried out by the human brain. In recent years, neural networks have become increasingly complex, with an ever-increasing number of parameters and energy consumption. While today's deep or convolutional neural networks perform excellently for many cognitive tasks, it isn't easy to use them to understand information processing in living organisms. However, some types of artificial neural networks remain more rooted in biology. These are the impulse neural networks (Spiking Neural Networks or SNN). They are based on communication mechanisms observed between neurons. The latter transmit information to each other through electrical impulses with reduced energy consumption. Learning within these SNNs can be carried out in an unsupervised way from plasticity rules observed in living organisms, such as Spike-Timing-Dependent Plasticity or STDP.

This thesis aims to model visual processing during locomotion from SNs regulated by STDP-type learning and which receive data captured by asynchronous cameras whose operation is directly inspired by the retina.

A first study ([Fricker et al., 2022]) looks at creating such a network for learning optical flow components. By interfacing this SNN with event data entries that were first simulated, we were able to highlight the network's performance for estimating the different components of the optical flow through different noise and speed conditions. The same network was later trained from actual navigation data captured using an event camera. While the task is more complicated despite convincing optical flow decoding, it highlights the need for a dataset halfway between simulated event data captured in actual conditions to get closer to natural navigation tasks while maintaining control over the data.

A second study focuses on creating such a dataset for network learning. This dataset is obtained from dedicated software to extract and then model the 3D structure within different environments. This software can then simulate the spikes generated by a virtual camera moving according to varying trajectories within these environments. This approach makes it possible to control the data received by the network and thus improve learning. The first results show that the neurons of the network become selective to the different components of the optical flow while remaining invariant to the spatial properties after learning. The proposed method, therefore, provides a solid basis for processing optical flow from impulse neural networks and within different environmental contexts.

Remerciements

J'aimerais tout d'abord remercier mes deux directeurs de thèse, Benoit Cottereau et Christophe Hurter, pour leur encadrement tout au long de cette thèse. Ils ont su m'épauler, m'écouter et me faire confiance pendant ces trois années. Cette thèse s'inscrivant dans un contexte multidisciplianire n'a pu que profiter de leur expertise mêlant à la fois neurosciences, intelligence artificielle ainsi que traitement et visualisation de données. J'ai pu grandir personnellement et professionnellement à leur côté et ont su me guider à travers les différentes étapes du doctorat.

Je remercie les directeurs successifs du laboratoire CerCo, Simon Thorpe et Isabelle Berry, ainsi que toute l'équipe du support informatique de m'avoir permis de travailler dans un cadre propice à la réflexion et la bonne conduite de mes travaux.

Je tiens à remercier les membres du laboratoire CerCo et de l'ENAC avec qui j'ai pu échanger au cours de mes travaux. Ainsi j'adresse un remerciement particulier à Tushar Chauhan avec qui j'ai pu longuement échanger et collaborer lors de son post-doctorat au CerCo. Sans son aide précieuse concernant les réseaux de neurones je n'aurai pu avoir une base aussi solide pour commencer ma thèse.

Je remercie également ceux avec qui j'ai passé la plus grande partie de mon temps au bureau, Guillaume Debat, Pauline Audurier et Mario Hervault. Nos échanges, sérieux ou non, ont toujours été un plus dont le bureau 130 se souviendra.

Cette thèse m'a également permis de rencontrer et d'échanger d'autres doctorants et ingénieurs au CerCo et à l'ENAC que je souhaiterais remercier pour leurs conseils et nos échanges. Ainsi ses remerciements vont à Ludovic

Gardy, Guillaume Truong et Augustin Degas.

Pour conclure je souhaite remercier ma famille pour leur soutien inconditionnel lors de ces trois dernières et plus largement lors de tous mes choix m'ayant conduit où j'en suis aujourd'hui. Je remercie également celle avec qui je partage mon quotidien qui a su me supporter et me soutenir dans les meilleurs jours comme dans les plus difficiles.

Finalement, ces derniers remerciements s'adressent à tous ceux que je n'ai pas cité mais qui méritent néanmoins ma reconnaissance.

Table des matières

Table des figures

Chapitre 1

Introduction

Lorsque nous marchons, courons ou plus simplement lorsque nous sommes en mouvement, notre système nerveux traite de nombreuses informations sensorielles qui nous permettent d'interagir avec notre environnement. Ces informations sont multiples et de différentes natures : sonores, haptiques ou encore visuelles. Leur intégration nous permet de prendre la décision adéquate une fois traitée (par exemple : est-il encore possible de traverser la rue sans prendre de risque ?). Parmi tous ces signaux, l'un prédomine les autres lors de la locomotion. En effet, le traitement de l'information visuelle joue un rôle majeur pendant la navigation et nous permet d'estimer nos trajectoires au travers d'environnements chargés, afin de suivre un chemin, d'estimer la distance parcourue ou encore d'éviter les collisions au sein de scènes dynamiques.

Si l'on s'intéresse aux phénomènes de perception visuelle et de détection des changements au sein de l'environnement, ceux-ci sont aujourd'hui mieux compris. Pour cela de nombreux travaux ont été nécessaires pour avancer dans ce domaine, depuis l'Antiquité jusqu'à notre ère. A de nombreuses reprises, philosophes et savants se sont heurtés à la définition de la vision et de ce qui fait que l'on voit avec nos yeux, ou plutôt selon Ptolémée (c.100-c.168) ce que l'on "touche" avec nos yeux. En effet, pour les philosophes de l'Antiquité, la perception visuelle relève de la perception de l'espace, la vision serait une forme de toucher qui se produirait par contact

sensoriel avec les objets qui nous entourent. Aussi Aristote (384-322 AEC) dans le second livre *De Anima : l'âme, les sens et les sensations* définit la vision comme l'actualisation de tout ce qui se trouve entre l'œil et l'objet observé et est permise par la projection des formes des objets dans l'œil. Euclide (c.300 AEC) quant à lui, émet la théorie des rayons visuels dans l'Optique, théorie selon laquelle des rayons émanant de l'œil entreraient en contact avec les objets observés au sein d'un cône, rapprochant également le sens de la vue à celui du toucher. Ces théories de la vision et du cône visuel, dites anciennes, vont se retrouver au coeur des études d'optique et de perception de la géométrie de l'espace jusqu'aux théories médiévales apportées par Ibn al-Haytham (c.965-c.1040), ou sous le nom latin Alhazen, dans le *Discours de la lumière* et le *Traité d'optique* [al Haytham et al., 1572]. Il est avancé que l'œil agit comme récepteur de l'information lumineuse et non pas comme émetteur. Ibn al-Haytham décrit alors la vision comme la déduction des propriétés distinctes de deux objets différents en percevant la taille, la forme, la couleur, la position et le mouvement. Ces avancées en permettent ainsi d'autres sur la compréhension de l'anatomie de l'œil et des processus cognitifs engagés lors de la perception visuelle qui seront alors la base de la découverte de l'image rétinienne par Johannes Kepler quelques siècles plus tard en pleine période de la Renaissance.

Le grand changement des façons de représenter le monde et de le percevoir ; c'est ainsi que pourraient être définis les courants de pensée des mathématiciens et philosophes de la Renaissance. Parmi eux Johannes Kepler (1571-1630), astronome et mathématicien, représente bien ce changement de la perception du monde en défendant l'hypothèse héliocentrique de Copernic par l'observation des astres, étoiles et planètes. C'est en étudiant l'orbite de Mars que Kepler s'intéresse à l'optique. En effet, le phénomène de réfraction de la lumière empêche la prise de mesures correctes et, cherchant à corriger ces dernières, Kepler étudie les travaux d'Ibn al-Haytham et rassemble ses connaissances et propres découvertes sur l'optique dans son livre *Astronomia pars Optica* et dans son traité *Dioptrica*. Dans ces derniers il présente les théories de la perspective et de la vision et décrit le phénomène de réfraction à travers l'atmosphère mais aussi les lentilles, étu-

diées notamment lors d'une éclipse solaire observée à l'aide d'une chambre noire. Cette étude de la chambre noire a ainsi permis d'étendre les concepts mathématiques impliqués aux mécanismes mis en jeu pour la perception visuelle et l'étude de l'œil. Si, à cette époque, le cristallin était considéré comme l'organe récepteur de lumière, Kepler avance alors qu'il n'est qu'une lentille permettant à l'image observée de se projeter sur la rétine qui la transmet au cerveau capable alors de la traiter correctement.

Depuis d'autres mathématiciens et philosophes se sont penchés sur la question de la perception visuelle comme René Descartes (1596-1650) qui écrira dans son traité *La Dioptrique*, complétant son *Discours de la méthode*, la loi des sinus corrigeant ainsi les équations de la réfraction de la lumière entre deux milieux selon Kepler. Il viendra également compléter le travail de Kepler sur l'optique physiologique et l'image rétinienne en y apportant ses travaux sur l'optique correctrice.

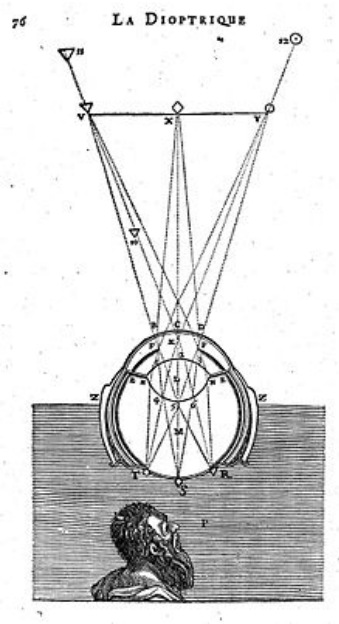


FIGURE 1.1 – Modèle de focalisation des images en diverses parties de la rétine initialement imaginé par Kepler et repris par Descartes dans son traité *La Dioptrique*.

Si l'on en revient à ce qu'est aujourd'hui la vision, le travail de Kepler a pu poser les bases du fonctionnement de l'organe sensoriel responsable de la vue, l'œil et plus particulièrement la rétine. Cependant, si ses travaux expliquent la composante sensorielle de la vision (i.e., la captation de la lumière au sein d'un environnement), les composantes de traitement et de mouvement restent toutes aussi importantes, ainsi que le traitement de l'information lumineuse reçue et comment ce traitement impacte nos prises de décision en fonction des informations intégrées. La compréhension des traitements s'effectuant au-delà de la captation de l'information lumineuse au sein de notre cerveau s'est grandement amé-

liorée depuis le siècle des Lumières au même titre que la compréhension du monde qui nous entoure. Ces traitements corticaux de données complexes et variées issues de l'observation de l'environnement ont pu être caractérisés et classés selon leurs natures et leurs actions sur l'information transmise. Ainsi cette dernière passe à travers un réseau cellulaire complexe de la rétine au cortex visuel.

En voyant ce réseau régissant les traitements effectués par l'humain pour l'intégration visuelle de son environnement, de nombreuses personnes se sont questionnées quant à la reproductibilité d'un tel réseau afin de pouvoir comprendre et modéliser les processus mis en jeu au sein du système nerveux. Ces questionnements ont alors donné naissance au domaine de l'intelligence artificielle puis à la création de réseaux dits artificiels cherchant à reproduire le fonctionnement et le comportement des réseaux corticaux biologiques. Ces modélisations se sont d'abord contenté de s'inspirer des processus de traitement biologiques, puis très vite les réseaux de neurones artificiels (ANNs) s'en sont affranchis pour établir des systèmes toujours plus performants. Ainsi depuis les premiers ANNs établis par [Rosenblatt, 1958], ces derniers ont subi bien des changements et leurs domaines d'applications se sont très fortement diversifiés. Leur architecture s'est complexifiée en multipliant les arrangements de neurones, des couches et de leurs connexions, faisant naître divers branches d'ANNs comme les réseaux de neurones convolutionnels (CNNs), les réseaux de neurones profonds (DNNs), ou même l'association des deux. Cette complexification des architectures et des modèles des ANNs s'est effectuée en fonction des besoins et des problématiques rencontrés dans les domaines du traitement d'image, du signal, des langues, etc. ou encore plus généralement pour des tâches d'optimisation, de simulation ou de classification.

Alors hyper efficaces, ces modèles souffrent néanmoins de limites car contrairement aux processus biologiques impliqués dans le cerveau, ces systèmes artificiels peuvent être très énergivores et très complexes, rendant leur compréhension difficile. Pour pallier cela, des systèmes ont vu le jour en alliant le principe de modélisation et reproductibilité du cerveau humain, tout en y intégrant des concepts et traitements biologiques. Ces systèmes

bio-inspirés constituent aujourd'hui une nouvelle génération de réseaux de neurones reposant sur le caractère impulsionnel de la transmission de l'information au sein du système nerveux ainsi que sur la grande adaptabilité du cerveau, notamment capable d'apprendre rapidement et sans supervision de nombreuses propriétés. Ces réseaux de neurones, définis par leur modèle de transmission impulsionnel et qualifiés de réseaux de neurones à impulsions, ou réseaux de neurones à spikes (SNNs) incarnent la troisième génération d'ANNs développés dans le domaine des neurosciences computationnelles et de la neuro-informatique [Maass, 1997]. Alliant les nouvelles technologies et les neurosciences, l'objectif de la neuro-informatique est d'aider à mieux comprendre le cerveau humain dans ses analyses et traitements notamment pour la vision et la locomotion. La modélisation des mécanismes d'apprentissage biologiques appliqués aux ANNs et à ses fonctions permet alors la mise en place de systèmes bio-inspirés que sont les SNNs afin d'ici traiter les indices visuels utiles à la locomotion.

1.1 Motivations

Cette thèse s'inscrit dans ce prisme des systèmes inspirés par la biologie. Au cœur du projet "Implants rétiniens pour l'aide à la Navigation Contextuelle chez les personnes Aveugles" (INCA, `cherchonspourvoir.org/projet-inca/`), dans le cadre d'une collaboration entre le laboratoire Centre de Recherche Cerveau et Cognition (CerCo, CNRS UMR 5549), l'Ecole Nationale de l'Aviation Civile (ENAC), l'Institut de Recherche en Informatique de Toulouse (IRIT) et la *National University of Singapore* (NUS), les travaux qui seront présentés ici invitent à se questionner sur le traitement de la vision chez l'humain à l'aide de systèmes bio-inspirés afin de pouvoir restituer les composantes s'y retrouvant dans un contexte de navigation. Ainsi les motivations de cette thèse sont les suivantes :

— La modélisation du traitement de l'information visuelle lors de la locomotion est majoritairement effectuée par des systèmes artificiels nécessitant une quantité importante de calculs entraînant une forte consommation énergétique aux antipodes des modèles biologiques

qui, associés à un grand nombre de paramètres, complexifient tout changement. La mise au point de systèmes bio-inspirés pour ce traitement des indices visuels pour la locomotion tout en se rapprochant du traitement humain permettra ainsi de mieux comprendre les mécanismes mis en jeu ainsi que leur apprentissage.

— L'optimisation des réseaux de neurones artificiels classiques repose sur la minimisation d'une fonction de coût définie mathématiquement et qui ne prend donc pas en compte les contraintes biologiques imposées par notre système nerveux comme son adaptabilité et sa plasticité. Par l'utilisation de réseaux de neurones à fonctionnement et apprentissage biologique plausible, les résultats que l'on pourrait obtenir seraient plus facilement interprétables et intégrables à des systèmes de restitution de l'information traitée.

Les systèmes bio-inspirés que je propose de développer seront constitués de réseaux de neurones à impulsions d'une part, dont le fonctionnement s'inspire directement du mode de transmission de l'information au sein du système nerveux chez le vivant, et de caméras événementielles d'autre part. Ces caméras ont la particularité de fonctionner de la même façon que la rétine humaine, tandis que le réseau de neurones à impulsions, ici doté de mécanismes d'apprentissage biologiques, représente l'étage de traitement des informations captées par la caméra. L'objectif de cette thèse est finalement de développer une approche qui ne souffre pas des limitations des systèmes artificiels classiques en élaborant de tels systèmes. Ceux-ci permettront une forte réduction des données à traiter, d'améliorer l'adaptabilité des réseaux grâce à une règle d'apprentissage non supervisée et de générer des sorties plus facilement interprétables.

Cette thèse a été financée par une bourse de l'Université Fédérale de Toulouse et de la région Occitanie délivrée dans le cadre du projet INCA après un stage de fin d'études que j'ai pu effectuer au CerCo autour des systèmes bio-inspirés pour la captation et le traitement d'informations visuelles. Ce stage a alors motivé la thèse afin d'explorer plus en détails le sujet étudié même sans une claire propédeutique en neurosciences ou en ANNs si ce n'est mon expérience mathématique, informatique et électronique. Dans

ce manuscrit seront alors retrouvés les travaux issus de la motivation apportée initialement, leurs développements et avancées tout au long de ces dernières années.

1.2 Plan du manuscrit

Ce manuscrit résume les principaux travaux qui ont rythmés la thèse au cours de ces dernières années. Le second chapitre s'intéresse aux mécanismes biologiques impliqués dans la perception visuelle et son traitement au sein du cortex visuel. Une attention particulière est apportée au traitement du mouvement et plus spécifiquement du flux optique au sein du système nerveux du primate humain mais aussi non humain (modèle macaque). Ce chapitre se clôture sur un état de l'art des modèles biologiques et computationnels pour le traitement du flux optique, ses mécanismes associés et son apport à la locomotion.

Le troisième chapitre décrit les systèmes bio-inspirés utilisés pour les travaux de la thèse. Il passe notamment en revue le fonctionnement des caméras événementielles, s'inspirant du traitement biologique effectué par la rétine. Sont également décrits les réseaux de neurones artificiels et leur développement au cours des dernières décennies pour en arriver aux réseaux de neurones impulsionnels, ce qui les composent et comment ils apprennent. Enfin, un état de l'art des études mettant en œuvre ce type de réseau de neurones clos ce chapitre, faisant écho au chapitre le précédant.

Les quatrième et cinquième chapitres décrivent les travaux réalisés au cours de la thèse, leurs intérêts, méthodes et résultats, ainsi que leur positionnement vis-à-vis de l'état de l'art défini en amont. Ainsi le chapitre 4 présente le réseau de neurones développé et maintenu tout au long de la thèse. Il est alors utilisé pour l'apprentissage de composantes de flux optique d'abord simulées, puis dans un contexte réel de navigation urbaine. Ces travaux ont fait l'objet de deux communications scientifiques lors de la conférence Bernstein 2021 Computational Neuroscience [Fricker et al., 2021] sous forme de poster, et lors de la conférence VISAPP 2022 pour un article de conférence [Fricker et al., 2022]. Le chapitre 5 reprend ce réseau établi

lors du précédent chapitre et est couplé à un jeu de données plus abouti pour son apprentissage. Le jeu de données créé permet l'exploration d'environnements réels capturés en 3D selon différentes trajectoires contrôlées par une méthode de lasergrammétrie. L'intégration de ces différents environnements s'effectue au sein d'un logiciel dédié et permet un déplacement libre de toute contrainte pour la génération d'événements compréhensibles pour le réseau de neurones.

Finalement, le chapitre 6, dernier chapitre de ce manuscrit, passe en revue les résultats obtenus, les perspectives et les enjeux futurs des travaux de cette thèse.

Chapitre 2

La perception visuelle, mécanismes et traitement du mouvement

Dans ce chapitre seront décrits les mécanismes de perception visuelle chez l'humain ainsi que le traitement de l'information visuelle par la rétine. Je détaillerai ensuite le fonctionnement du cortex visuel, son organisation et ses différentes voies, notamment la voie responsable de la perception et du traitement des informations de mouvement et de direction, ainsi que de la perception du flux optique. Finalement seront décrits différents modèles biologiques et computationnels des mécanismes de flux optique pour des tâches de perception visuelle et de locomotion.

2.1 Le système visuel humain : de la rétine au cortex visuel

La rétine

La rétine est située sur la surface interne du globe oculaire. Les traitements qu'elle effectue permettent de transformer les rayons lumineux reçus en un signal nerveux qui sera transmis au reste du système visuel. Elle

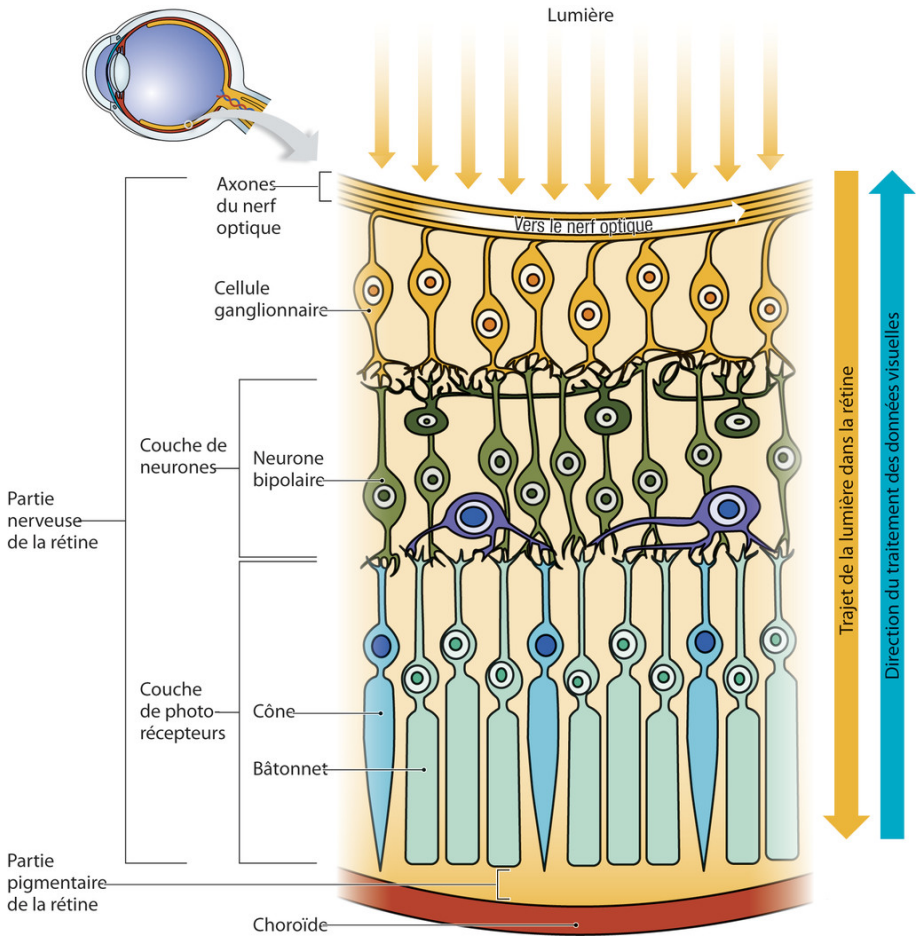


FIGURE 2.1 – Anatomie de la rétine, adapté de [Tortora and Derrickson, 2016].

consiste en une fine membrane d'un tissu neurosensoriel réagissant aux stimulations visuelles extérieures. Les rayons lumineux constituant cette information visuelle doivent d'abord traverser l'œil, par la cornée, le cristallin et l'humeur vitrée avant de se projeter sur les photorécepteurs constituant la première couche de la rétine.

Ces photorécepteurs peuvent être classés en deux catégories distinctes, présents sur deux zones rétiniennes différentes. Tout d'abord la macula, zone centrale de la rétine comptant pour 5 degrés du champ de vision en son centre, se retrouve avec une population majoritaire de cônes : photorécepteurs permettant la vision diurne (avec leur concentration maximale sur

la zone de la fovéa comptant pour 1 degré du champ de vision). Sur le reste de la rétine, en périphérie de la macula, les bâtonnets - photorécepteurs permettant la vision nocturne - sont majoritaires. L'information visuelle une fois réceptionnée par cette couche de photorécepteurs traverse alors la couche nucléaire interne composée des cellules bipolaires, amacrines et horizontales, puis la couche des cellules ganglionnaires avant d'être acheminée vers le nerf optique via les axones des cellules ganglionnaires.

Ainsi les cellules bipolaires décrivent la voie de transmission directe de l'influx nerveux émanant des photorécepteurs après la captation de l'information visuelle et vers les cellules ganglionnaires, tandis que les cellules amacrines et horizontales vont venir moduler le signal nerveux et permettre une adaptation du contraste, des couleurs ou encore une meilleure appréciation des contours grâce à l'intervention des cellules horizontales en amont [Kuffler, 1953].

Les cellules horizontales, considérées comme interneurones puisque présentes entre les photorécepteurs et les cellules bipolaires, jouent un rôle de rétroaction sur les photorécepteurs [Herrmann et al., 2011]. Ces cellules inhibent de manière sélective l'information transmise par les photorécepteurs vers les cellules bipolaires en fonction de l'intensité lumineuse perçue par zones rétiniennes locales. Cela se traduit alors par un mécanisme de contrôle de gain local permettant le maintien de l'information visuelle en signal nerveux dans une plage de fonctionnement adéquate aux circuits rétiniens internes [Verweij et al., 2003, Wässle, 2004].

D'autres interneurones interviennent après les cellules bipolaires et avant les cellules ganglionnaires, les cellules amacrines. Se connectant alors aux cellules bipolaires, ganglionnaires et entre elles, leurs sorties constituent la principale entrée des cellules ganglionnaires [Jacoby et al., 1996]. Les cellules amacrines identifiées chez le rongeur comme jouant un rôle dans la détection du mouvement d'un objet [Olveczky et al., 2003] et dans la sélectivité à la direction de mouvement [Hausselt et al., 2007] en influant sur les réponses des cellules bipolaires afin de les transmettre aux cellules ganglionnaires, leurs rôles chez le primate restent encore à définir précisément [Diamond, 2017]. En effet, les cellules amacrines constituent la classe de

cellules rétiniennes la plus vaste, chacune ayant des rôles variés et parfois multiples. Bien que l'on sache aujourd'hui qu'elles agissent sur les réponses lumineuses des cellules bipolaires et ganglionnaires, l'identification de leurs connexions synaptiques et ainsi de leurs rôles restent à définir [Grünert and Martin, 2020].

Les cellules ganglionnaires constituent la dernière étape du traitement de l'information visuelle dans la rétine avant son envoi vers le cortex occipital via le nerf optique. Celles-ci vont transformer le signal analogique qui leur est transmis en signal électrique. Ce signal électrique se traduit par la génération de potentiels d'action comme réponses à l'information initialement reçue [Stone, 2012]. Ces réponses sont alors organisées par champs récepteurs correspondant aux champs visuels responsables de l'émission de potentiels d'actions : chaque photorécepteur réagit à une portion spatiale du champ visuel total définissant son propre champ récepteur qui, combiné à d'autres en passant par les cellules bipolaires vont former à leur tour leur champs récepteurs, pour finalement constituer les champs récepteurs des cellules ganglionnaires à partir de ceux des cellules bipolaires. Ces champs récepteurs se retrouvent être organisés de manière antagoniste par leur polarisation, présentant deux régions concentriques, un centre réagissant aux variations de luminosité positives dit centre 'ON', et un pourtour réagissant aux variations de luminosité négatives dit pourtour 'OFF', ou inversement avec un centre 'OFF' et un pourtour 'ON'. Ce type de structure des champs récepteurs permet alors de détecter les contrastes au sein d'une scène visuelle et d'en réduire le bruit [Petkov and Subramanian, 2007]. Si les cellules ganglionnaires sont différentiables par leurs champs récepteurs, elles le sont aussi par leurs types. Trois types de cellules ganglionnaires sont aujourd'hui identifiées, les cellules naines, parasols et bistratifiées [Dacey, 2000, Dacey, 2004, Koch et al., 2004, Nassi and Callaway, 2009]. Les cellules naines grâce à leur haute résolution spatiale, leur faible sensibilité aux basses fréquences temporelles et leur taille de champ récepteur réduite, se voient être en charge de la vision fine. Les cellules parasols vont à l'inverse être responsables de la vision grossière par leur faible résolution spatiale et leur sensibilité aux hautes fréquences temporelles.

Ces différents types de cellules ganglionnaires vont alors également définir différentes voies du traitement visuel : les voies parvocellulaire, magnocellulaire et koniocellulaire, dont le traitement se voit être séparé au sein du corps géniculé latéral (CGL), relié à la rétine par le nerf optique, et du cortex visuel.

Le corps géniculé latéral

Le CGL consiste en une structure considérée comme relais des informations visuelles en provenance de la rétine passant par le nerf optique. Ces afférences provenant de la rétine sont arrangées de manière rétinotopique : l'organisation spatiale des informations visuelles reste la même que celle ayant été projetée sur la rétine. Ainsi des neurones situées les uns à côté des autres au sein du CGL répondent à des zones proches au sein du champ visuel.

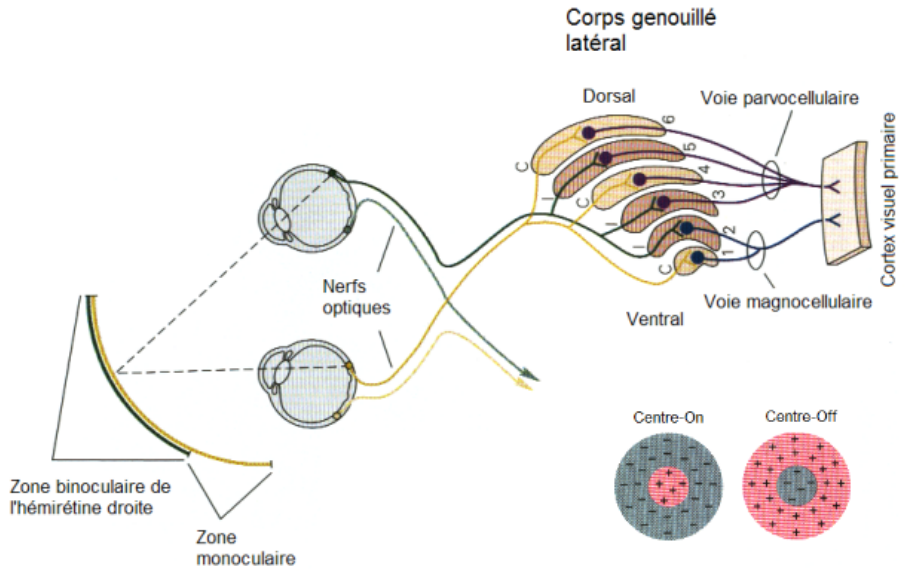


FIGURE 2.2 – Le CGL, ses différentes couches reliées entre la rétine et le cortex visuel primaire, et ses champs récepteurs de centre-ON et centre-OFF, adapté de [Pidoux, 2011].

Le CGL est décomposable en six couches. Les deux couches ventrales dites couches magnocellulaires reçoivent les informations provenant de la voie magnocellulaire, tandis que les quatre couches dorsales appelées couches parvocellulaires reçoivent les informations provenant de la voie parvocellulaire [Briggs and Usrey, 2011]. Aussi les afférents de la voie koniocellulaire se voient être reçus par les neurones koniocellulaires intercalés entre les différentes couches.

Les voies parvocellulaire et magnocellulaire vont alors relayer des informations spécifiques vers le cortex visuel. Les neurones constituant les deux différentes couches du CGL font apparaître des propriétés différentes, principalement leur sensibilité aux différents contrastes de couleurs. Ainsi les neurones parvocellulaires répondent aux changements de couleur mais peu aux changements de luminosité. A l'inverse, les neurones magnocellulaires répondent peu aux changements de couleurs mais sont sensibles aux changements de luminosité. Les champs récepteurs du CGL reflètent quant à eux le même antagonisme centre-ON / pourtour-OFF (ou inversement) observé au sein des cellules ganglionnaires tout en présentant une taille qui varie en fonction de l'excentricité rétinienne [Crook et al., 1988]. Le CGL vient aussi jouer un rôle de filtrage et relaie l'information visuelle par un filtrage passe-bas [Heiberg et al., 2013].

Finalement, les sorties du CGL conduites par les radiations optiques sont projetées au sein du cortex visuel primaire (V1) afin de traiter l'information reçue.

Le cortex visuel

Les signaux afférents du CGL viennent alors se projeter dans le cortex visuel, et notamment au sein de l'aire V1, première aire du système visuel. Le système visuel est hiérarchique et se base sur un ensemble d'aires responsables du traitement de l'information visuelle qui se distinguent par leur propriétés fonctionnelles, leurs connectivités ainsi que leurs organisations rétinotopiques. Au fur et à mesure de l'avancée de l'information au sein des différentes régions fonctionnelles du cortex visuel, les propriétés

visuelles traitées sont de plus en plus complexes. Ici sont succinctement décrites les aires visuelles rétinotopiques médiales-postérieures (V1, V2, V3) ainsi que leurs principaux efférents (V4 et V5) selon le modèle hiérarchique [Van Essen and Maunsell, 1983, Felleman and Van Essen, 1991] :

- Le cortex visuel primaire V1 reçoit les informations provenant du CGL et en projette sur d'autres structures cérébrales. Son organisation et ses fonctions seront décrites en détail dans la prochaine section.
- Le cortex visuel secondaire V2, distinguable par la présence de trois types de bandes : fines, épaisses et intermédiaires. Chacune de ces bandes est responsable de la sensibilité aux couleurs pour les bandes fines, à l'orientation et la direction de mouvement pour les bandes épaisses. Recevant les informations de V1, elles sont réparties vers les aires supérieures selon leurs caractéristiques [Zeki, 1978, Felleman and Van Essen, 1991, Van Essen et al., 2001].
- L'aire V3 reçoit les informations visuelles venant des bandes épaisses de V2 ainsi que de V1. Les neurones composant V3 ont une haute sensibilité au contraste et se voient être sélectifs à l'orientation et au mouvement [Tootell et al., 1997, Zeki, 2003].
- L'aire V4 se retrouve impliquée dans le traitement de la couleur ainsi que des formes en fonction de leurs couleurs aidant à l'identification des objets de la scène visuelle [MEADOWS, 1974, Lueck et al., 1989, Hadjikhani et al., 1998].
- L'aire V5 ou MT (*Middle Temporal*) est une aire où la grande majorité des cellules est sélective au mouvement et présente une préférence de direction ou d'orientation permettant ainsi le traitement des objets en mouvement ainsi que la sélectivité aux directions de mouvement [Kolster et al., 2010, Snowden et al., 1992, Van Essen et al., 1981]. L'aire MT est directement connectée à l'aire MST (*Medial Superior Temporal*) codant le flux optique et ses différentes composantes. Ces aires seront détaillées par la suite.

Il est important de noter que le traitement de l'information visuelle n'est pas seulement hiérarchique car les neurones au sein des différentes

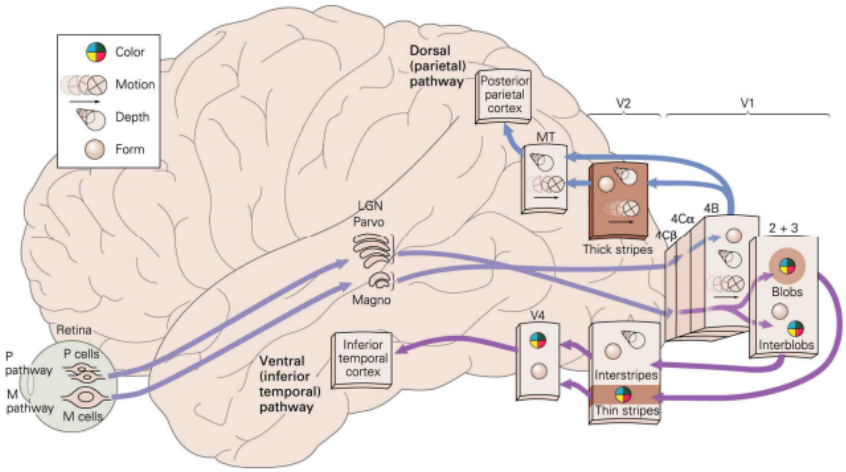


FIGURE 2.3 – Les différentes aires du cortex visuel, leur organisation ainsi que les différentes fonctions cognitives auxquelles elles sont associées. Les voies ventrale (ou voie du Quoi) et dorsale (ou voie du Où) sont respectivement représentées en violet et en bleue. Figure adaptée de [Behnke, 2003].

aires reçoivent aussi des signaux d'autres neurones au sein de la même aire (via des connexions cortico-corticales) et aussi de neurones situés dans des aires de plus haut niveau (via des connexions feedback) [Maunsell and van Essen, 1983, Kennedy and Bullier, 1985, Boyd and Casagrande, 1999]. De plus, l'organisation de ces aires corticales révèlent l'existence de deux voies de traitement appelées voies dorsale et ventrale.

Dans la section suivante, il est décrit en détail les fonctions réalisées par l'aire visuelle primaire, V1.

Le cortex visuel primaire

L'aire V1, aussi appelée cortex visuel primaire ou cortex strié, située dans la partie postérieure du cerveau, reçoit les informations provenant du CGL et plus particulièrement les informations émanant des trois voies magnocellulaire, parvocellulaire et koniocellulaire. Ces voies viennent alors se projeter dans V1 au sein de différentes couches et notamment dans la

couche 4C. Cette couche 4C reçoit la grande majorité des informations visuelles afférentes du CGL et se sous-divise en deux couches, les couches $4C\alpha$ et $4C\beta$. Les trois voies viennent alors se projeter sur ces différentes couches où la détection des mouvements, des formes et des couleurs se retrouve séparée.

Les afférents de la voie magnocellulaire vont se projeter sur la couche $4C\alpha$ de V1, ceux de la voie parvocellulaire sur la couche $4C\beta$ et ceux de la voie koniocellulaire sur des blobs au sein des couches 1, 2, 3 et 4A. Il est alors retrouvé tout le long de la voie rétino-géniculo-striée cette discrimination entre ces trois différentes voies. Cette organisation horizontale en couches fait intervenir différents types de neurones pour la réception et l'envoi de l'information à laquelle s'ajoute une organisation verticale dite en colonnes où les différents neurones réagissent aux même caractéristiques d'une région donnée du champ visuel définissant ainsi les champs récepteurs de V1. Les champs récepteurs de V1, issus d'une convergence de plusieurs champs récepteurs de cellules antérieures de type centre-pourtour, se trouvent alors être plus allongés et leurs propriétés dépendent de la couche dans laquelle ils se situent et du type de neurone associé. Ces types de neurones sont de deux catégories : les cellules simples et les cellules complexes [Hubel and Wiesel, 1968].

Les cellules simples sont principalement présentes dans la couche 4 de l'aire V1 et se manifestent par un champ récepteur allongé et divisé en deux à trois parties excitatrices ('ON') ou inhibitrices ('OFF'). Le profil spatial de ces champs récepteurs peut alors être décrit par une fonction Gabor à centre 'ON' et périphérie 'OFF', ou inversement, permettant la détection de l'orientation et de la position de barres. Ce profil particulier de champ récepteur s'explique par les connexions entre les neurones de V1 à plusieurs cellules ganglionnaires via le CGL, représenté par la figure 2.4.

Les cellules complexes quant à elles se trouvent au sein des couches 2, 3, 5 et 6 de V1. Répondant aussi à l'orientation spécifique de barres, elles se différencient des cellules simples par un champ récepteur uniforme dépourvu de zones ON ou OFF bien définies. Ces cellules se retrouvent à répondre de façon optimale au déplacement d'une barre à une orientation

précise quel que soit son emplacement au sein du champ récepteur là où les cellules simples ont besoin d'un signal fixe et correctement orienté. Aussi, de la même façon que les champs récepteurs des cellules simples sont formés à partir des cellules ganglionnaires du CGL, les champs récepteurs des cellules complexes sont formés à partir des réponses des cellules simples [Hubel and Wiesel, 1962].

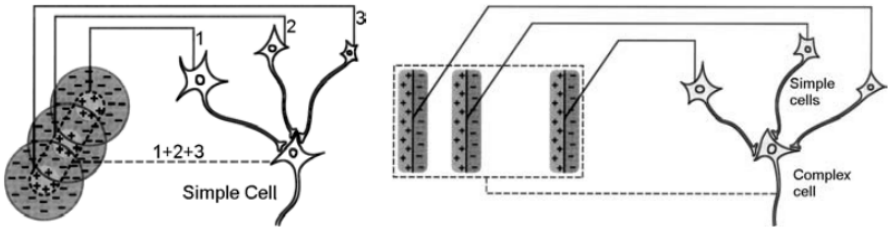


FIGURE 2.4 – Les cellules simples et complexes. Selon [Hubel and Wiesel, 1968], les cellules simples combinent les réponses des cellules ganglionnaires du LGN selon leurs champs récepteurs concentriques pour former leurs propres champs récepteurs allongés. Ces champs récepteurs sont sensibles à l'orientation et la phase des stimuli. Les réponses de plusieurs cellules simples présentant une orientation similaire, mais une sélectivité à la phase différente sont combinées pour former les champs récepteurs des cellules complexes. Ces champs récepteurs montrent alors une réponse invariante à la phase à des barres ou des angles orientés (tiré de [Behnke, 2003]).

Il est alors aussi décrit une troisième catégorie de cellules par [Hubel and Wiesel, 1965], les cellules hypercomplexes. Ces cellules présentent les mêmes caractéristiques que les cellules complexes, hormis la présence bien délimitée de deux zones, 'ON' et 'OFF', au sein de leurs champs récepteurs. Si elles répondent également à l'orientation de barres, elles prennent aussi en compte leur longueur et leur largeur. Ainsi plus la barre en mouvement reste dans la zone 'ON', plus la réponse de la cellule associée est élevée, et diminue au fur et à mesure que la barre se déplace dans la zone 'OFF'. On retrouvera alors les différents profils spatiaux des champs récepteurs des différents types de cellules ainsi que leurs activités en fonction du stimulus dans la figure 2.5.

Comme évoqué précédemment, à une organisation horizontale en couches

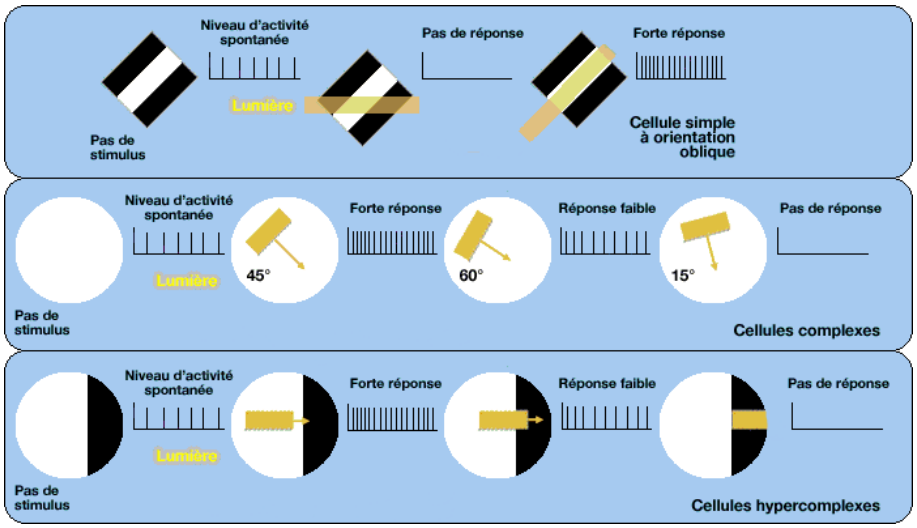


FIGURE 2.5 – Activité des différents types de cellules selon la présentation d'un stimulus visuel au sein de leurs champs récepteurs. De haut en bas : Le champ récepteur d'une cellule simple sélectif à une orientation oblique, le champ récepteur d'une cellule complexe sélectif à une orientation à 45 degrés, le champ récepteur d'une cellule hypercomplexe réceptif à une orientation horizontale et à une longueur de barre équivalente aux deux tiers de la taille du champ récepteur (adapté de [Beaubert et al., 2005]).

de V1 s'ajoute une organisation verticale en colonnes. Ainsi les cellules présentant une préférence similaire à l'orientation sont regroupées dans une structure en colonne, cette colonne partageant alors une même sélectivité à l'orientation et des champs récepteurs couvrant une même zone du champ visuel [Hubel et al., 1977]. La répétition et l'organisation de ces différentes colonnes suivent un cycle régulier de la sélectivité à l'orientation, illustré par les couleurs dans la figure 2.6. Au sein de ces colonnes se retrouvent les régions de blobs au sein des couches 2 et 3 qui répondent aux couleurs et aux basses fréquences spatiales [Livingstone and Hubel, 1984]. Ainsi, toutes les colonnes étant sélectives au cycle complet des orientations possibles avec les régions de blobs forment un ensemble appelé hyper colonne, occupant un millimètre carré du cortex, pour une zone précise du champ visuel total, amenant les sélectivités aux différentes orientation, couleurs, et un traitement stéréoscopique par l'alternance des colonnes de dominance oculaire.

Ces colonnes d'orientation et de dominance oculaire sont représentées par la figure 2.6.

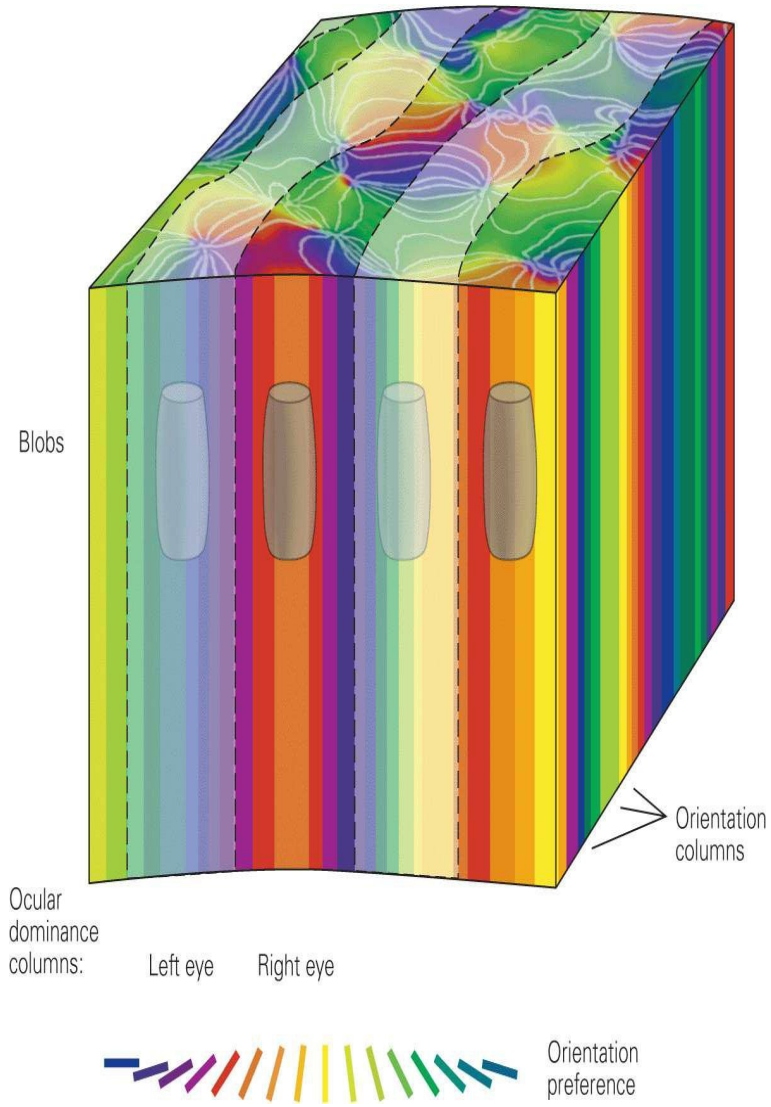


FIGURE 2.6 – Les deux structures en colonnes au sein de l'aire V1 formant les structures d'hyper-colonnes. Ici sont représentées les colonnes d'orientations en couleurs correspondantes à leur sélectivité à l'orientation, ainsi que les colonnes de dominance oculaire correspondant aux champs visuels de l'œil droit ou gauche, tiré de [Kandel et al., 2012].

V1 joue un rôle clé pour le traitement des informations visuelles. Cette aire est impliquée dans la détection des orientations, des contrastes et égale-

ment des directions de mouvement en effectuant un filtrage spatio-temporel de l'information venant du CGL et de la rétine. Si l'information visuelle est traitée par V1, elle en sort également et est transmise par la suite aux différentes aires V2, V3, V4 et MT notamment. Comme nous l'avons vu plus haut, il est alors possible de distinguer deux chemins empruntés par l'information visuelle, en charge de traitements différents : la voie ventrale et la voie dorsale.

Les différentes voies de traitement

Le système visuel se sépare alors en deux voies distinctes pour le traitement des information visuelles [Mishkin et al., 1983]. Il a été proposé que ces deux voies soient séparées en fonction de leur rôle : la voie ventrale traite les informations concernant l'identification et la reconnaissance des objets grâce à leurs différentes caractéristiques, tandis que la voie dorsale traite les informations sur l'emplacement et le mouvement des objets de la scène. La voie ventrale est aussi appelée la voie du "Quoi ?", et la voie dorsale, la voie du "Où ?" [Ungerleider and Mishkin, 1982].

Bien que largement adopté depuis des décennies, ce modèle se voit être remis en question par de récents arguments émanant de différentes études [Sheth and Young, 2016, Rossetti et al., 2017, Pitcher and Ungerleider, 2021]. Les arguments de la remise en question de cette claire ségrégation entre les voies du "Quoi ?" et du "Où ?" se portent sur la responsabilité non exclusive de la voie ventrale pour la reconnaissance d'objets d'une part, et selon quoi la voie dorsale ne serait pas la seule responsable de la vision spatiale et du contrôle visuo-moteur mais interviendrait également au sein des mécanismes d'attention visuelle et spatiale d'autre part.

Cette différenciation des voies du cortex visuel n'est alors pas aussi définie et précise qu'elle pourrait laisser paraître dû à l'organisation spatiale des zones corticales mises en jeu : réparties dans le cortex pariétal pour la voie dorsale, et dans le cortex temporal pour la voie ventrale. Des interconnexions existent alors entre les différentes zones et la distinction des voies ne peut être absolue à travers toute la chaîne du traitement visuel. Toujours

est-il que des mécanismes restent propres à ces deux voies distinctes et que leurs rôles en tant que voie du "Quoi ?" et voie du "Où ?" restent largement admis [Ungerleider and Mishkin, 1982, Goodale and Milner, 1992, Norman, 2002].

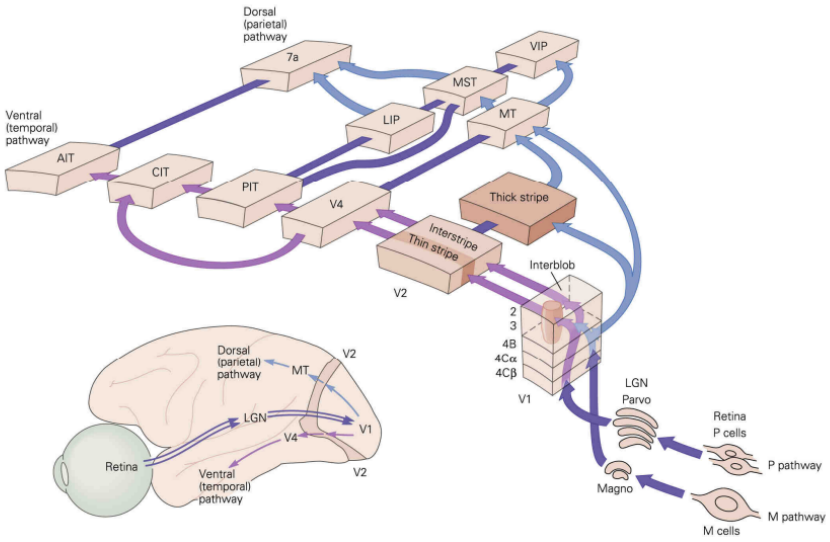


FIGURE 2.7 – L'organisation des voies visuelles dites dorsale et ventrale au sein du cortex visuel. La voie dorsale, allant de V1 au cortex pariétal, est associée au traitement spatial de la vision et du mouvement (voie du "Où ?"). La voie ventrale, allant de V1 au cortex inférotemporal, sert à la reconnaissance d'objets et des formes (voie du "Quoi ?"), tiré de [Behnke, 2003].

La voie dorsale est très importante dans le cadre de ce travail de thèse car c'est elle qui est impliquée dans le traitement des données spatiales et de la perception du mouvement. Dans la section suivante, je reviens plus en détails sur les aires corticales qui composent cette voie, et notamment sur leur fonctionnement ainsi que sur leur implication dans la perception spatiale lors de la locomotion.

2.2 La voie dorsale : perception et traitement du mouvement

La voie dorsale comprend plusieurs régions corticales dont l'aire MT, l'aire MST et les aires intrapariétales ventrale et latérale (VIP et LIP). Les informations reçues et traitées par ces aires proviennent directement d'aires de plus bas niveau et notamment des aires V1 et V2.

Comme évoqué précédemment, les neurones magnocellulaires de la rétine et du CGL sensibles à une faible luminance, aux faibles fréquences spatiales et aux fortes fréquences temporelles apportent les informations relatives au traitement du mouvement dans la voie dorsale. Celles-ci se projettent au sein de la couche $4C\alpha$ de l'aire V1, où sont présentes les cellules complexes sensibles au mouvement de barres et contours orientés [Hubel and Wiesel, 1968, Hubel et al., 1978, Adelson and Bergen, 1985], ainsi qu'à la direction de mouvement [Orban et al., 1986, Movshon and Newsome, 1996]. Les cellules de l'aire V1 répondent essentiellement aux mouvements locaux (i.e., au sein de leur champ récepteur) au sein de patterns plus complexes [Movshon and Newsome, 1996]. Une sélectivité précoce à la vitesse est observable dans les réponses des cellules de l'aire V1 par la combinaison des fréquences spatio-temporelles des cellules complexes [Orban et al., 1986, Priebe et al., 2006]. L'information de mouvement est ensuite projetée vers les bandes épaisses de l'aire V2 depuis la couche 4B de l'aire V1 [Livingstone and Hubel, 1987, Levitt et al., 1994]. Ainsi les bandes épaisses de l'aire V2 transmettent leurs informations de traitement de mouvement vers l'aire MT [DeYoe and Van Essen, 1985, Shipp and Zeki, 1985, Born and Bradley, 2005].

Si l'aire MT est alors la prochaine étape du traitement du mouvement après l'aire V2, l'aire MT reçoit aussi des informations directement depuis l'aire V1 où le traitement du mouvement se fait à une échelle locale, tandis qu'à travers l'aire MT se retrouve un traitement plus global du mouvement [Felleman and Van Essen, 1991, Born and Bradley, 2005]. Les cellules présentes dans l'aire MT sont sensibles à différentes composantes associées au mouvement en deux dimensions telles que la direction, la vitesse et la fré-

quence spatiale [Maunsell and van Essen, 1983, Lagae et al., 1993, Priebe et al., 2003, Brooks et al., 2011]. Les champs récepteurs des neurones de l'aire MT sont beaucoup plus grands et plus étendus que ceux de l'aire V1 et permettent ainsi de traiter des mouvements plus globaux [Pack and Born, 2001, Britten et al., 1992, Snowden et al., 1992]. Cela permet finalement aux neurones de l'aire MT de traiter le mouvement de plusieurs points ou contours à la fois émanant d'un même ou de multiples objets en mouvement, ou encore de résoudre le problème de l'ouverture grâce à un traitement du mouvement unidirectionnel uniquement sur une partie du champ visuel [Pack and Born, 2001].

Ainsi, l'aire MT se retrouve impliquée dans le traitement du mouvement mais également, et dans une moindre échelle, impliquée dans la perception de mouvement "simples" tel que le mouvement unidirectionnel. Suivant la hiérarchie du système visuel, les aires après MT sont impliquées dans des traitements de formes plus complexes de mouvement, comme le traitement du flux optique dans l'aire MST notamment. La description de ces traitements fait l'objet de la prochaine section.

Flux optique

Le flux optique désigne la projection des mouvements des objets et des environnements qui nous entourent sur nos rétines lors de nos déplacements. Son traitement par le système nerveux donne accès à des informations sur la structure de la scène observée et aussi utiles pour la navigation comme la direction de déplacement ou la distance parcourue. Ce flux optique peut se décomposer en différentes composantes en fonction du type de déplacement effectué. Lors d'un déplacement vers l'avant ou vers l'arrière, le flux optique est décrit suivant ses composantes radiales, l'expansion et la contraction. Tous les points de l'espace se déplacent alors vers l'extérieur du champ visuel allant de la vision centrale vers la vision périphérique lors de l'expansion / déplacement vers l'avant, et inversement pour la contraction / déplacement vers l'arrière. Lors de balancements et de rotation de la tête ou de déplacements en translation dans l'espace, d'autres composantes se

manifestent, les composantes rotationnelles ou translationnelles.

Ces différentes composantes de mouvements traduisent un mouvement global de l'observateur dans son environnement grâce aux indices visuels et lumineux projetés sur la rétine. Ces composantes vont alors inclure les trois patterns décrits plus haut (translationnelles, rotationnelles et radiales), tous observables lors de différentes conditions de déplacement. Les composantes translationnelles se manifestent lors de mouvements de translation dans l'espace de la droite vers la gauche et inversement, ainsi que lors de déplacements du haut vers le bas et inversement comme lors de sauts ou d'accroupissements par exemple. Les composantes rotationnelles peuvent se manifester lors des mouvements de balancement de la tête lors de la locomotion par exemple, ainsi amener son oreille vers son épaule induira une composante rotationnelle dans le sens horaire ou anti-horaire. Les composantes radiales, comme présentées précédemment, se manifestent lors de déplacements vers l'avant ou vers l'arrière induisant des patterns d'expansion ou de contraction.

Le flux optique joue alors un rôle dans le comportement d'un observateur vis à vis de ses déplacements et sert plusieurs fonctions. [Gibson, 1950] décrit dans son livre T*he Perception of the Visual World* le flux optique comme étant le pattern de mouvement rétinien des éléments présent dans l'environnement généré lorsque l'on se déplace dans cet environnement. Un observateur peut ainsi percevoir grâce au flux optique la direction de son mouvement propre [Gibson, 1947, Gibson, 1950, Warren, 1976, Warren et al., 1988]. Le flux optique fournit aussi des indices visuels permettant diverses actions comme pouvoir se diriger vers une cible définie aussi appelé *heading* [Calvert, 1954, Gibson, 1958, Warren et al., 2001], aider au maintien de l'équilibre en plus de la proprioception [Lee and Aronson, 1974, Stoffregen, 1985], identifier l'environnement et les objets, leurs formes et leurs dispositions [Koenderink, 1986, Todd, 1995, Rogers, 2021], ou encore d'éviter ou de provoquer des collisions en sachant estimer la vitesse et la position de ce qui nous entoure par rapport à notre propre vitesse et position dans l'espace, appelé *flow parsing* [Lee, 1976, Warren and Rushton, 2007, Warren and Rushton, 2008, Warren and Rushton, 2009].

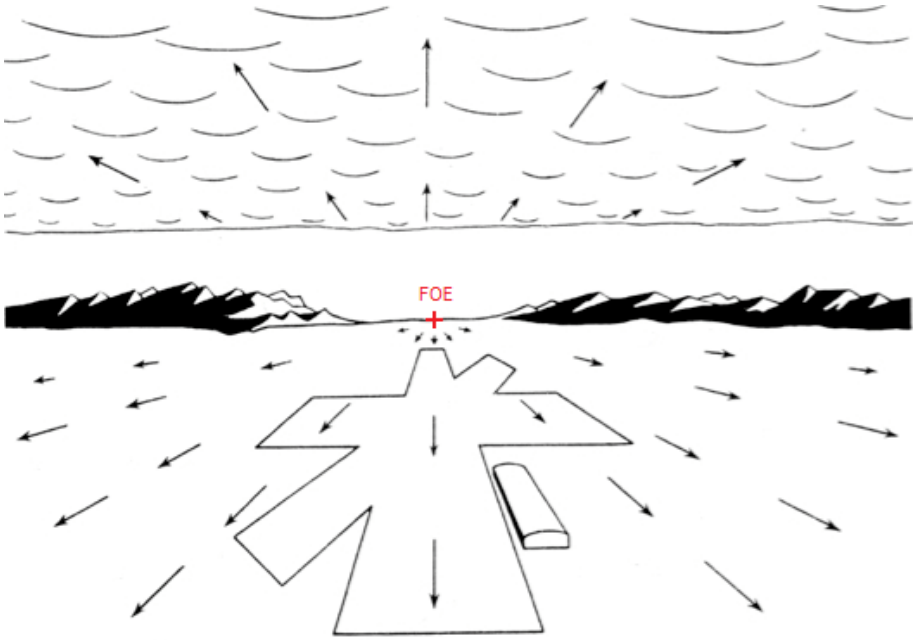


FIGURE 2.8 – Le flux optique, représenté par les flèches, comme perçu lors d'un atterrissage par avion sur piste. Les composantes du flux présentent un pattern radial, une expansion vers l'avant nous indiquant nous diriger vers le point d'expansion (FOE), adapté de [Gibson, 1950].

Si l'on s'intéresse au traitement de ce flux optique, ses composantes et les différents mécanismes mis en jeu au sein de la voie dorsale, ce dernier est possible grâce à l'information du mouvement local et global fournie par l'aire MT, transmise à l'aire MST. L'aire MST se retrouve alors impliquée dans le traitement du mouvement en trois dimensions, ainsi que le début du traitement du flux optique et du mouvement de soi. Divisée en deux régions, latérale (MSTl) et dorsale (MSTd), l'aire MSTl traite alors les informations de vitesse des trajectoires des objets en mouvement pour la poursuite oculaire [Tanaka and Farah, 1993, Ilg, 2008]. Les neurones de l'aire MSTd quant à eux sont sélectifs aux composantes de rotation, d'expansion-contraction, et de leur combinaison [Graziano and Gross, 1994, Takahashi et al., 2007, Mineault et al., 2012]. Les neurones de l'aire MSTd se retrouvent finalement sélectifs au flux optique [Duffy and Wurtz, 1991] et peuvent en déduire le *heading* ou le mouvement propre [Duffy and Wurtz, 1995, Gu

et al., 2006].

Heading

Le flux optique décrit par [Gibson, 1950] peut informer sur la direction du mouvement lors de la navigation, son "cap", son *heading* [Bruss and Horn, 1983, Longuet-Higgins and Prazdny, 1980] puisque lors d'un déplacement linéaire vers l'avant, sans mouvements parasites des yeux ou de la tête, le point d'origine des patterns de flux optique communément appelé point de fixation ou point d'expansion (ou FOE pour *focus of expansion*) vient s'aligner avec le *heading* de l'observateur. [Warren et al., 1988, Cutting et al., 1992] ont montré la capacité d'un observateur à déterminer son *heading* à l'aide des indices du flux optique et ainsi pouvoir effectuer des tâches de locomotion sans encombre. De plus l'estimation du *heading* depuis le flux optique se révèle être résistante aux perturbations extérieures propres à la locomotion humaine telles que les balancements et les rebonds du corps et de la tête, ou la perturbation du champ de vision par des composantes de mouvements parasites [van den Berg, 1992, Kim et al., 1996, Cutting and Readinger, 2002].

Malgré la robustesse de cette estimation du *heading*, cette dernière est altérée par un autre type de mouvement naturellement présent au sein du champ visuel lors de la locomotion. Le mouvement généré par un objet se déplaçant indépendamment de l'observateur peut alors biaiser une estimation correcte. Par exemple, des objets obstruant le FOE et translatant face à un observateur en mouvement vont générer un biais de l'estimation de *heading* corrélé à la direction de mouvement des objets [Royden and Hildreth, 1996, Li et al., 2018]. Tandis que si ces objets au lieu de rester à une distance fixe, se rapprochent de l'observateur, induisent un biais de l'estimation de *heading* dans la direction opposée des objets en mouvements [Warren and Saunders, 1995, Li et al., 2018].

Si ces biais d'estimation peuvent être causés par des objets en mouvement, cela est dû au système visuel qui, bien que capable de correctement identifier les mouvements de chaque objet, vient associer tous les mouve-

ments perçus dans le champ visuel pour l'estimation du *heading* [Li et al., 2018, Riddell et al., 2019]. Entre alors en jeu un second mécanisme issu de l'analyse du flux optique et permettant l'estimation de la vitesse et la position des objets par rapport à la vitesse propre et la position propre d'un observateur. C'est finalement après l'aire MST, où la voie dorsale continue vers les aires intrapariétales que le traitement du mouvement inclut un traitement plus complexe des composantes du flux optique et notamment le *flow parsing* [Phinney and Siegel, 1999, Raffi and Siegel, 2007, Chen et al., 2013, Raffi et al., 2013].

Flow parsing

Comme exposé précédemment, le flux optique apporte les informations de *heading* ainsi que les informations visuelles et de mouvements de l'environnement. Il a été également décrit bien que le système visuel humain permet à un observateur d'estimer son *heading*, cette estimation se voit être biaisée par l'éventuelle présence d'objets animés dans la scène observée causant des divergences d'estimation de quelques degrés d'angle visuel [Royden and Hildreth, 1996, Warren and Saunders, 1995, Li et al., 2018].

Cependant, pouvoir estimer la vitesse des objets qui nous entourent tout en nous déplaçant peut s'avérer crucial dans des conditions nécessitant l'évitement des collisions, par exemple lors de déplacements au sein d'une foule mobile ou pendant la conduite au sein d'un environnement où le trafic est dense. Également, correctement estimer notre propre vitesse en fonction d'un objet en mouvement nous permet de l'attraper, comme récupérer une balle lancée vers nous par exemple. Alors, afin de pouvoir effectuer correctement ces tâches, notre mouvement propre doit être compensé afin de faire basculer les coordonnées d'un objet en mouvement centrées sur la rétine, égocentré, vers des coordonnées centrées sur l'environnement extérieur, ou allocentré. Le déplacement d'un objet se voit donc être caractérisé par ses coordonnées sur la rétine, résultant des coordonnées de l'objet au sein de l'environnement dans lequel il se déplace et du flux optique induit par le mouvement propre de l'observateur de l'objet en mouvement. Ainsi, le mé-

canisme de *flow parsing* se voit être l'évaluation du mouvement d'un objet dans son environnement auquel le système visuel vient soustraire le flux optique généré par le mouvement propre de l'observateur au mouvement de l'objet projeté sur la rétine [Warren and Rushton, 2007, Warren and Rushton, 2008, Warren and Rushton, 2009].

Pour illustrer le mécanisme de *flow parsing*, prenons l'exemple donné par la figure 2.9. Un observateur se déplace vers l'avant produisant un flux optique présentant un pattern radial d'expansion. Un objet en mouvement entre dans le champ visuel et vient alors produire un mouvement sur la rétine résultant de la combinaison de son mouvement dans son environnement et du flux optique radial provoqué par le mouvement propre de l'observateur. Afin d'évaluer le déplacement de l'objet par rapport à l'environnement, le mécanisme de *flow parsing* propose que le système visuel compense le flux optique induit par le mouvement propre, et dans cet exemple vient rajouter la composante inverse de l'expansion, la contraction à chaque partie en mouvement de l'image incluant le mouvement de l'objet observé. Le *flow parsing* permet alors de percevoir le mouvement propre d'un objet se déplaçant dans son environnement par compensation des composantes de flux optique générées par un observateur en mouvement à l'emplacement de l'objet observé [Foulkes et al., 2013, Fajen et al., 2013, Dokka et al., 2015].

Néanmoins, si le traitement du flux optique et de ses différentes composantes est utile pour la locomotion, l'établissement de cartes spatiales et l'estimation du trajet parcouru, d'autres mécanismes sont également nécessaires. Ainsi la locomotion se base aussi sur d'autres modalités sensorielles comme la proprioception. Certaines études associées à ce traitement multisensoriel sont présentées au cours de la prochaine section.

Traitement du flux optique à travers la voie dorsale pour la locomotion

Pour se déplacer au sein d'environnements dynamiques et complexes, l'humain se fie à sa perception de mouvement, son propre déplacement ainsi que le déplacement des objets environnants grâce aux informations extraites

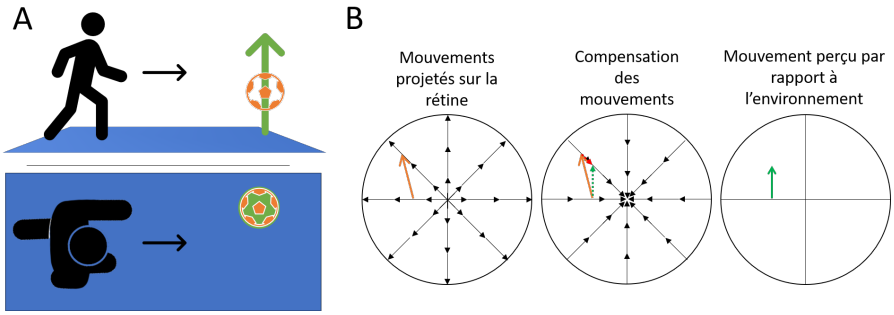


FIGURE 2.9 – Le mécanisme de *flow parsing* ici illustré lors d'un déplacement vers l'avant avec observation d'un objet en mouvement. A) Déplacement d'un observateur dans un environnement vu de côté et de dessus. L'observateur se déplace vers l'avant et observe un objet se déplaçant vers le haut en face à gauche de lui. B) Les différentes étapes composant le mécanisme de *flow parsing*. Les mouvements projetés sur la rétine sont ceux de la composante radiale du flux optique décrivant un déplacement vers l'avant de l'observateur illustrée par les flèches noires, et le mouvement de l'objet illustré par la flèche orange. Parce que l'observateur se déplace vers l'objet en mouvement, le mouvement de ce dernier projeté sur la rétine de l'observateur est décrit par une composante verticale et une légère composante translationnelle vers la gauche. Le mécanisme de *flow parsing* vient ensuite compenser toutes ces différentes composantes de mouvement et vient donc compenser la composante translationnelle vers la gauche de l'objet avec une composante translationnelle vers la droite. Finalement, après compensation, il ne reste plus que le mouvement propre de l'objet dans son environnement représenté par la flèche verte. (Adapté de [Peltier et al., 2020]).

à partir du flux optique. Ainsi ces informations permettent une locomotion assurée pouvant à la fois servir à maintenir un *heading* tout en prenant en compte les informations de *flow parsing* permettant l'identification, la détection et l'estimation des mouvements d'objets pendant la locomotion. Ici seront décrits les différents modèles, issus d'études électrophysiologiques et psychophysiques chez le primate, mais aussi computationnelles permettant de mettre en lumière les mécanismes sous-tendant le traitement du flux optique.

Les modèles biologiques chez le primate

Au cours de ces dernières années, plusieurs groupes de recherche se sont intéressés aux différents mécanismes du flux optique et ses composantes à travers son traitement dans la voie dorsale et pendant le déplacement de soi.

[Sato et al., 2010] ont pu mettre en évidence que les propriétés spatio-temporelles des réponses de l'aire MST combinent le mouvement des objets et les composantes du flux optique afin de représenter le mouvement de soi à travers différentes conditions naturalistiques. Pour ce faire, ils ont enregistré les réponses des neurones de l'aire MSTd chez le primate non-humain (modèle macaque) à des stimuli de flux optique et d'objets en mouvement afin de simuler les indices de mouvement propre durant la translation. Ces stimuli présentaient différentes configurations naturelles de composantes du flux optique ainsi qu'une superposition d'objets animés, se déplaçant de manière indépendante ou fixés au sol. Ces deux types d'objets induisent les mêmes réponses au niveau cortical en interaction avec le flux optique dont certaines reflètent des sensibilités au mouvement local et global. La sensibilité au mouvement local est alors basée sur l'arrangement spatial des directions créées par l'objet animé et le flux optique qui l'entoure, tandis que la sélectivité au mouvement global se base sur la fréquence temporelle du *heading* préféré, traduisant un chemin emprunté au sein de l'environnement. Il est finalement mis en évidence que les réponses spatio-temporelles des neurones de l'aire MST répondent aux composantes de flux optique combinées aux mouvements des objets environnants afin d'en déduire le mouvement propre et le chemin emprunté grâce au *heading* au cours de tâches de locomotion.

Les populations de neurones au sein de l'aire MST intègrent des indices visuels au cours de la locomotion, notamment le flux optique, mais cette intégration peut également être multisensorielle. [Takahashi et al., 2007, Gu et al., 2010] montrent alors que les réponses de l'aire MST intègrent à la fois le flux optique et le mouvement inertiel donné par les informations vestibulaires. L'intégration de ces différentes informations se voit être utile lors

de la locomotion au sein de divers environnements pour des tâches de *heading* mais également de rotation en trois dimensions. En effet, la figure 2.10 met en évidence l'intégration des données visuelles et vestibulaires par les neurones de l'aire MST afin de déterminer le mouvement propre. Cela met également en évidence des neurones dits congruents, comme pour la condition de translation, où les taux d'activation moyens reflètent une même tendance spatiale, et des neurones opposés qui, à l'inverse et comme observable pour la condition de rotation, montrent des profils spatiaux opposés.

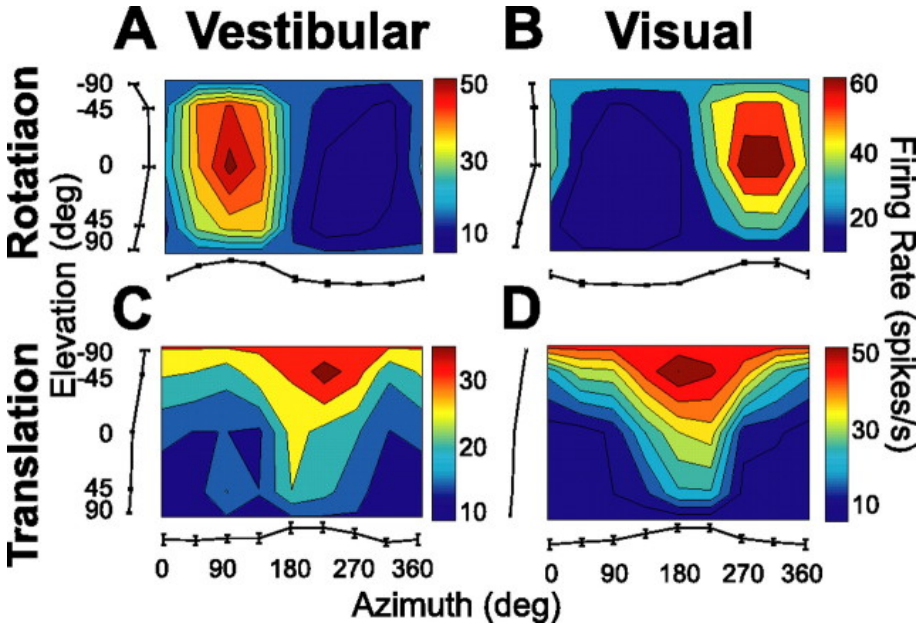


FIGURE 2.10 – Réponses d'un neurone de l'aire MSTd testé durant une rotation vestibulaire (A), rotation par stimulation visuelle grâce au flux optique (B), translation (*heading*) vestibulaire (C), translation (*heading*) par stimulation visuelle grâce au flux optique (D). Le code couleur met en évidence le taux d'activation moyen en fonction des angles d'élévation et de l'azimut de l'observateur, tiré de [Takahashi et al., 2007].

Les résultats suggèrent alors que les neurones congruents contribuent à une perception robuste des indices de navigation pour la perception du *heading*, condition lors de laquelle les informations visuelles et vestibulaires montrent les mêmes composantes de mouvement. Quant aux neurones opposés, puisque répondant lorsque les informations visuelles et vestibulaires

sont contradictoires, ceux-ci joueraient alors un rôle pour la détection d'objets se déplaçant indépendamment du mouvement induit par notre propre déplacement. Finalement, les cellules dites opposées seraient importantes pour la décomposition du mouvement projeté sur la rétine, discriminer le mouvement des objets du mouvement propre, pour le *flow parsing*.

Plusieurs groupes de recherche se sont également penchés sur la question du traitement du *flow parsing* au sein des mécanismes du flux optique traité dans l'aire MST. [Warren and Rushton, 2009] d'abord, s'intéressent à mettre en exergue le rôle important du traitement du flux optique pour la détection et l'estimation d'objet en mouvement lors de la locomotion en soustrayant alors les composantes de mouvement propre aux mouvement de la scène visuelle perçus, décrivant le mécanisme de *flow parsing*.

Leur première expérience consiste en un observateur voyant projeté devant lui une simulation de flux optique suivant une composante radiale d'expansion constituée d'un nuage de points. Parmi ces points se trouve une cible se déplaçant de façon indépendante dont l'observateur doit retranscrire la trajectoire à l'aide d'un témoin linéaire à orienter selon la trajectoire perçue de cette cible en mouvement. Tout ceci est représenté par la figure 2.11. A cela s'ajoute trois conditions (voir figure 2.11-C) : le nuage de points simulant le flux optique apparaît entièrement à l'observateur ("Full"), le flux optique n'apparaît qu'autour de la cible en mouvement à évaluer("Local"), et le flux optique est présent partout sauf où la cible est présente ("Global"). Ces différentes conditions permettent alors de mesurer l'impact des indices de mouvements locaux et globaux. Les données recueillies grâce à cette première expérience permettent aux auteurs de montrer que la trajectoire perçue d'un objet en mouvement est largement plus influencée par des mécanismes de perception globale du flux optique, tel que le *flow parsing*, avec une contribution beaucoup plus faible d'un traitement local.

Pour aller plus loin, dans une seconde expérience est proposé de tester la présence effective d'un traitement global en ne retirant pas uniquement une petite partie du flux optique, mais en occultant l'intégralité d'un hémichamp visuel. Trois différentes conditions sont alors également testées (voir figure 2.11-D) : le flux optique est complètement présent ainsi que la

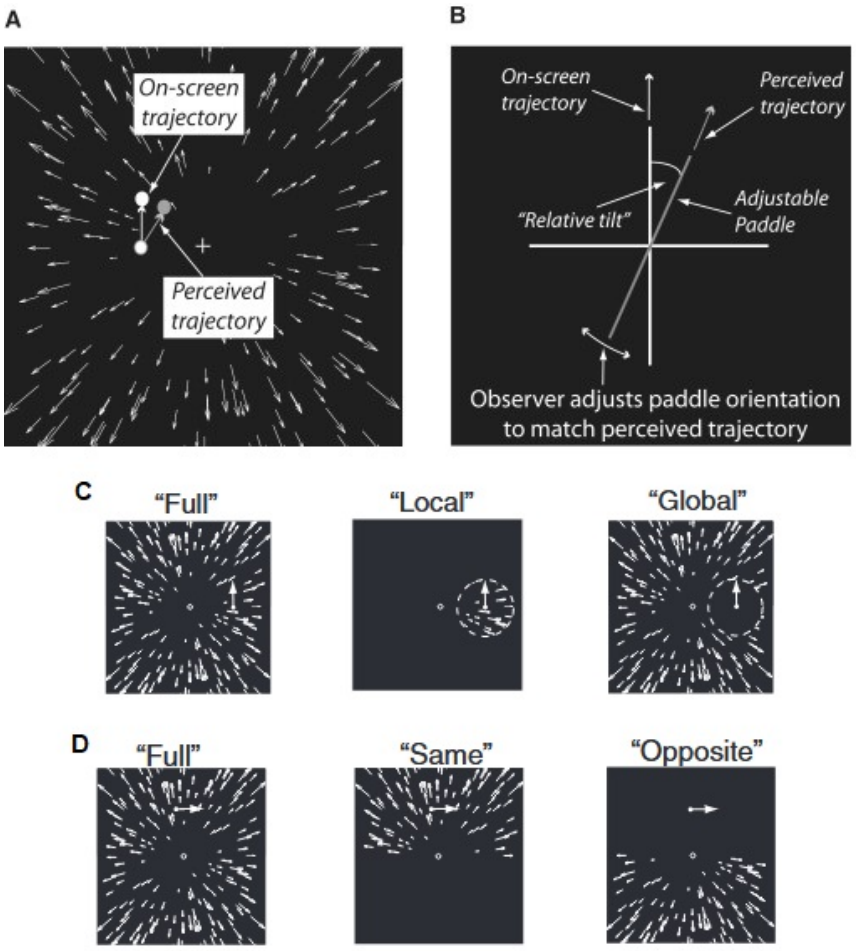


FIGURE 2.11 – (A) Simulation du flux optique utilisé par [Warren and Rushton, 2009]. Elle décrit un nuage de point suivant une composante d'expansion du champ visuel à laquelle s'ajoute une cible se déplaçant de manière indépendante. (B) La tâche de l'observateur est alors de définir le mouvement perçu de la cible à l'aide d'un témoin à incliner. (C-D) Les différentes conditions des deux expériences, tiré de [Warren and Rushton, 2009].

cible est en mouvement ("Full"), l'hémichamp où la cible est présente est conservé et l'autre est supprimé ("Same"), l'hémichamp où la cible est présente est supprimé et l'autre est conservé ("Opposite"). La tâche à effectuer reste la même, traduire le mouvement perçu de la cible à l'aide d'un témoin à incliner. Les résultats de cette expérience montrent alors que le phénomène de *flow parsing* peut intervenir alors même que la cible à observer est

spatialement isolée du flux optique.

Finalement, [Warren and Rushton, 2009] concluent sur l'existence d'un mécanisme purement visuel et se reposant sur un traitement global des informations visuelles permettant d'identifier et de soustraire le mouvement perçu dû au propre déplacement d'un observateur, son mouvement propre, le mouvement de soi. Ce mécanisme de perception globale de la scène observé pour l'appréhension des mouvements des objets lors de nos propres mouvements, identifié comme *flow parsing*, montre alors la capacité du cerveau à intégrer cette solution pour la locomotion à l'aide du flux optique.

L'intégration du *flow parsing*, mécanisme alors défini précédemment comme purement visuel, ne serait alors pas le seul contributeur pour la locomotion et la perception du déplacement des objets lors du mouvement propre. Des indices non-visuels entrent aussi en jeu comme il est avancé par [Fajen et al., 2013]. Ainsi les auteurs montrent que si le mécanisme de *flow parsing* permet effectivement à un observateur en mouvement de soustraire ses propres composantes de déplacement au sein du flux optique perçu par rapport aux objets en mouvement dans la scène, l'aide à la locomotion procurée par ces indices visuels - estimations de vitesse et de direction investiguées dans cette étude - est complémentaire aux indices non-visuels. Ce raisonnement s'appuie sur de précédents travaux tels que ceux de [Tcheang et al., 2005, Dyde and Harris, 2008] où le jugement d'un objet en mouvement se trouvait plus précis lorsque les informations de mouvement propre étaient effectivement présentes. [Fajen et al., 2013] appuient sur le possible rôle de ces combinaisons de mécanismes visuels tel que le *flow parsing* et l'estimation d'indices visuels grâce au flux optique généré par le mouvement propre, complétés par les informations non-visuelles, comme les informations vestibulaires avancées précédemment dans la descriptions des études de [Takahashi et al., 2007, Gu et al., 2010], permettent d'apporter une aide à la locomotion pour des tâches d'interception ou d'évitement d'objets en mouvement.

Se pose alors la question de la reproductibilité de tels mécanismes. Ainsi pouvoir retrouver les résultats de ces études électrophysiologiques et psychophysiques au travers de modèles mathématiques et computationnels per-

mettrait de mieux comprendre le fonctionnement du traitement du flux optique pour la locomotion au sein des aires visuelles dédiées chez le primate, à savoir les aires MT et MST.

Les modèles computationnels inspirés par la biologie

[Beyeler et al., 2016] se basent sur des mesures effectuées en électrophysiologie chez le macaque [Takahashi et al., 2007] pour modéliser les réponses de l'aire MST aux composantes du flux optique lors de la locomotion et notamment au cours de tâches de *heading*. Le modèle ainsi mis au point permet de mieux comprendre les mécanismes permettant l'émergence de la sélectivité au mouvement et au flux optique au sein des aires MT et MST. [Beyeler et al., 2016] proposent un modèle se reposant sur une technique appelée factorisation de matrice non négative (NMF, voir figure 2.13). Cette NMF permet de retrouver, à partir de réponses simulées en sortie d'un mo-

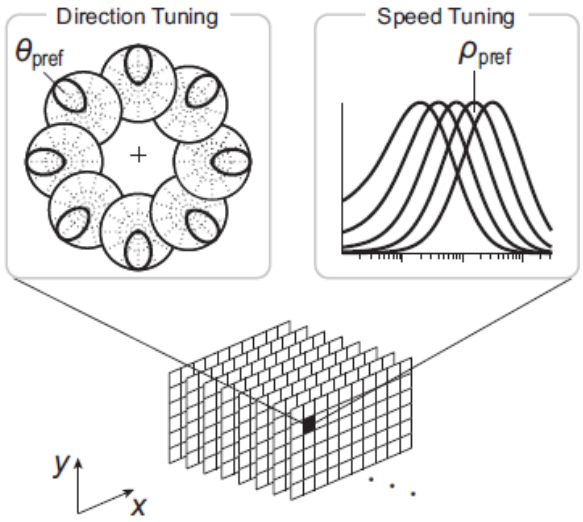


FIGURE 2.12 – Modèle de l'aire MT utilisé par [Beyeler et al., 2016] pour la simulation en sortie du modèle de réponses de l'aire MT à des champs de flux optique causé par le déplacement d'un observateur. Chaque champ de flux est alors traité par des unités de MT modélisées, chacune étant paramétrée selon une direction θ_{pref} et une vitesse spécifique ρ_{pref}.

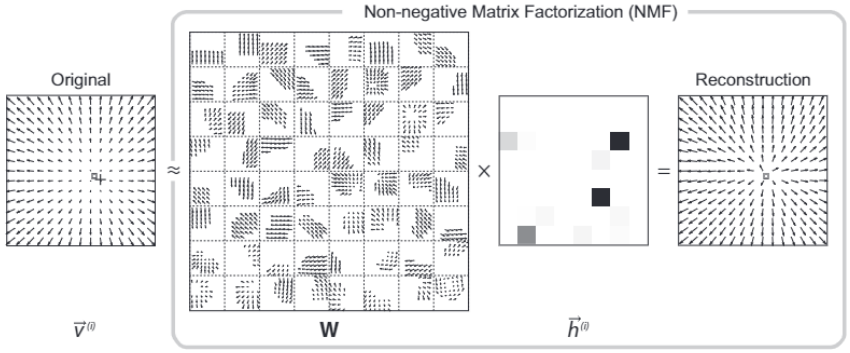


FIGURE 2.13 – Le principe d'application de la NMF d'après [Beyeler et al., 2016]. La NMF appliquée à des sorties simulées de neurones de l'aire MT induit une représentation éparse du flux optique observé et de ses différentes composantes à travers différents champs récepteurs. Ces champs récepteurs et leurs champs vectoriels associés représentant le flux optique observé (la matrice W) sont alors multipliés par leurs coefficients associés dans la matrice vectorielle h afin que leur produit se rapproche le plus possible de la matrice vectorielle v. Le but de l'algorithme de NMF est finalement de trouver une décomposition de la matrice de données v, avec la contrainte que tous les éléments des matrices W et h soient non négatifs.

dèle de l'aire MT (décrit par la figure 2.12), les sorties des neurones de l'aire MST en lui présentant des stimuli de rotation ou de *heading*. Les auteurs comparent alors leurs résultats obtenus par NMF à ceux obtenus chez le vivant par [Takahashi et al., 2007] et ces derniers se montrent cohérent par rapport au modèle biologique de l'aire MST comme le montre la figure 2.14. Cette modélisation computationnelle permet alors de comprendre un peu mieux les mécanismes en jeu au sein de l'aire MSTd pour la sélectivité au mouvement et au flux optique, et que ces sélectivités à des mouvements complexes peuvent être réduites à de simples factorisations de différentes composantes pour différentes fonctions induites par le flux optique comme le *heading*.

L'étude du mouvement de soi et du *heading* à travers des modèles computationnels depuis des modèles biologiques s'est aussi vu traiter par d'autres groupes de recherche. En effet, [Steinmetz et al., 2022] établissent

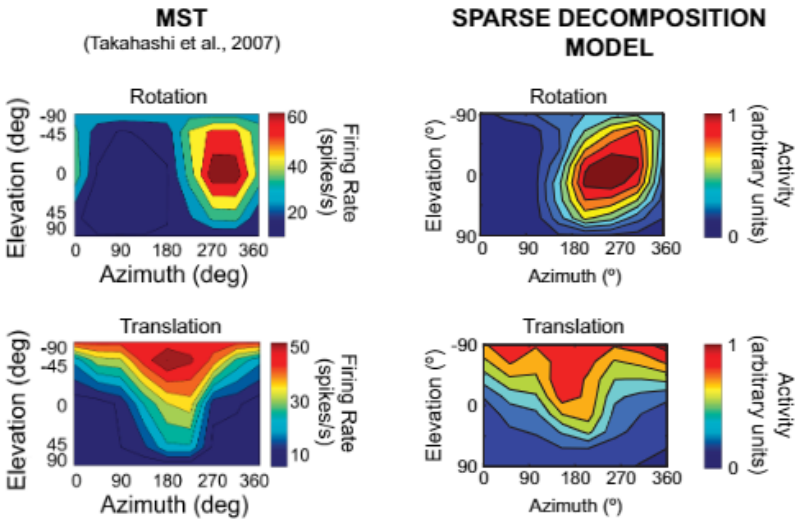


FIGURE 2.14 – Comparaison des résultats obtenus en électrophysiologie par [Takahashi et al., 2007] et en modélisation par [Beyeler et al., 2016]. Les préférences d'orientation des neurones de l'aire MSTd dans les conditions de rotation et de translation (ici comme *heading*) partagent les mêmes taux d'activation pour ces deux modèles validant ainsi le modèle computationnel vis à vis du modèle biologique.

un modèle computationnel selon le principe d'encodage sensoriel efficace à partir d'une méthode mise au point par [Ganguli and Simoncelli, 2014]. Ce principe d'encodage sensoriel efficace, lui-même tiré d'une hypothèse biologique de codage efficace initialement décrite par [Barlow, 1961], propose que les communications entre les neurones des systèmes sensoriels forment un codage neuronal commun pour efficacement représenter et communiquer les informations sensorielles. Ainsi ce code permet un échange d'informations demandant une charge de transmission minimale entre chaque neurone. L'encodage sensoriel efficace rendu dynamique (DESE pour *Dynamic Efficient Sensory Encoding*) par [Steinmetz et al., 2022] se trouve être un mécanisme potentiellement biologique (voir par exemple [Brenner et al., 2000] pour une illustration à partir de mesures chez la mouche bleue de la viande *Calliphora vicina*) en tant que mécanisme de réglage neuronal s'adaptant aux stimuli sensoriels. L'utilisation du DESE permet alors dans

un modèle neuronal que décrivent [Steinmetz et al., 2022], et illustré par la figure 2.15, de paramétrer les unités le composant afin que grâce à des stimuli visuels, il soit possible de détecter le flux optique et la vitesse des composantes qui le compose. Ce modèle est alors capable de produire des estimations de la direction du mouvement propre basés sur le flux optique.

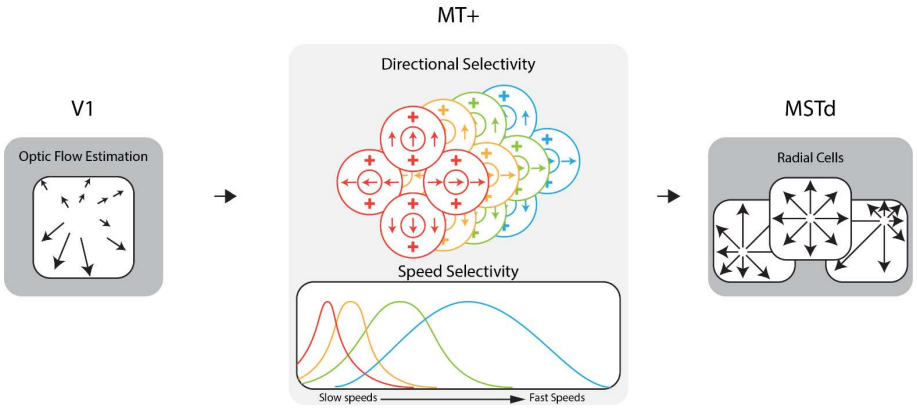


FIGURE 2.15 – Diagramme décrivant le modèle de [Steinmetz et al., 2022]. En entrée se trouve le flux optique tel qu'il serait perçu dans l'aire V1 avant d'être transmis à l'aire MT+ qui encode le signal du flux optique reçu à travers ses différents champs récepteurs qui sont chacun paramétré pour une certaine vitesse de détection et une position donnée. Finalement les cellules modélisées de l'aire MT+ transmettent leurs données aux cellules de l'aire MSTd qui sont elles sensibles aux différentes composantes du flux optique. L'utilisation du mécanisme de DESE intervient dans le paramétrage des courbes de sélectivité à la vitesse des cellules simulées de l'aire MT+.

Cette aide à la locomotion grâce au *flow parsing* s'est aussi vue être appuyée par [Layton et al., 2012] au travers d'un modèle computationnel permettant l'estimation du *heading* d'un observateur en présence d'objets se déplaçant indépendamment au sein du champ visuel observé. En répliquant les simulations de flux optique utilisés dans les études psychophysiques menées par [Royden and Hildreth, 1996] et [Warren and Saunders, 1995], les auteurs font observer ces simulations non pas à des participants humains mais à un modèle computationnel s'inspirant du système visuel humain à différents étages de traitement apparentés aux aires V1, MT+ et MSTd. Ces simulations de flux optiques sont décrites comme des nuages de point décri-

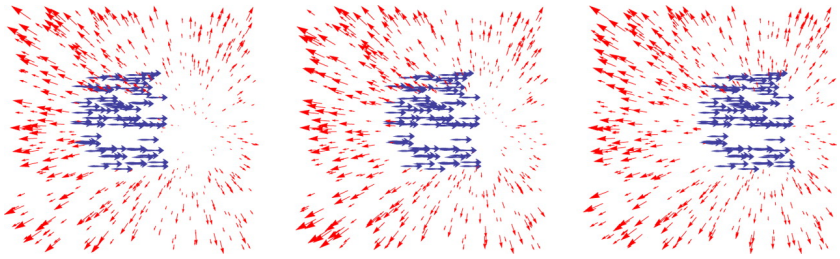


FIGURE 2.16 – Simulation du flux optique selon la composante radiale d'expansion (en rouge) utilisée par [Warren and Saunders, 1995, Royden and Hildreth, 1996, Layton et al., 2012] pour l'estimation du *heading* lorsqu'un objet obstrue le FoE de l'observateur par translation (en bleu). Ici est représenté le champ vectoriel du nuage de point utilisé pour la représentation de la direction emprunté par les différents points simulés.

vant une composante radiale du flux optique au sein desquels un groupe de points se déplace indépendamment du reste de la scène visuel, induisant le déplacement d'un objet. Le champ vectoriel de ces simulations est illustré par la figure 2.16. En faisant observer ces simulations à des participants humains afin qu'ils déterminent le *heading* perçu, [Warren and Saunders, 1995, Royden and Hildreth, 1996] s'aperçoivent qu'un biais de détermination du *heading* est induit par l'objet en mouvement et que ce biais est corrélé à la direction de déplacement de l'objet lorsque celui-ci passe devant le FoE, s'approchant ou non de l'observateur. Le modèle utilisé par [Layton et al., 2012] observant ces mêmes simulations de flux optique fait apparaître ces mêmes biais dans l'estimation du *heading* lors de la comparaison des résultats obtenus entre le modèle et les études psychophysique (voir figure 2.17). Le modèle computationnel indique alors que le système visuel humain est capable de déterminer le *heading* grâce aux traitements du mouvement et du flux optique dans les aires MT et MST sans le besoin de neurone sensible au mouvement différentiel. Au contraire, la distribution des neurones de l'aire MST dans le modèle computationnel indique une adaptation au flux optique présenté par leurs temps de réponse variant lorsque l'objet en mouvement est présent ou non (voir figure 2.17). Les auteurs concluent alors que leur modèle démontre que les biais observés chez l'humain lors de l'estimation du *heading*, accompagné de mécanismes de *flow parsing* par

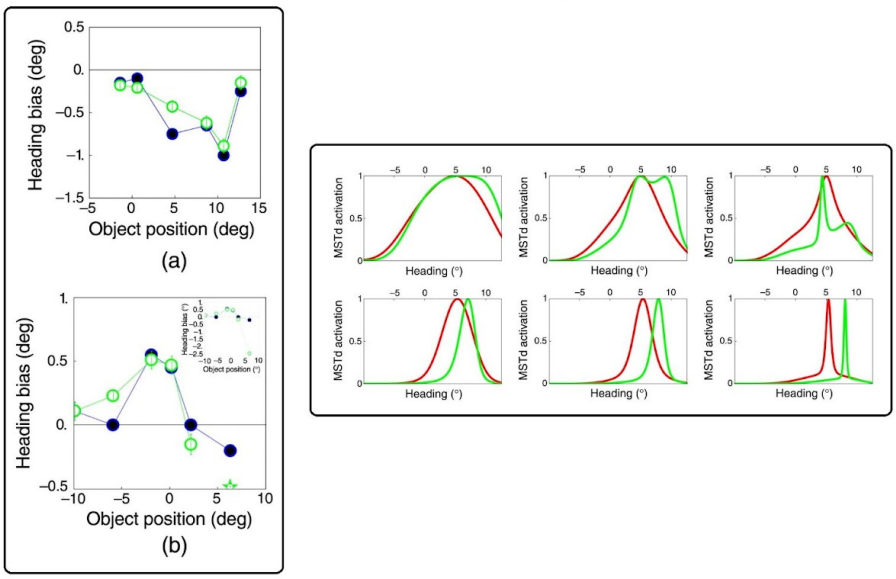


FIGURE 2.17 – Résultats obtenus par [Layton et al., 2012]. Panneau de gauche : Ici est présenté le biais dans l'estimation du *heading* par rapport à la position de l'objet en mouvement, vers la gauche (a) et vers la droite (b), au sein du champ visuel avec en vert les résultats du modèle et en bleu les résultats des participants humains obtenus par [Royden and Hildreth, 1996]. Panneau de droite : Les réponses des neurones de l'aire MSTd durant différentes présentations de simulations en nuage de points sans la présence d'un objet en mouvement (en rouge) et avec (en vert). Les graphiques de la ligne du haut représentent les activations à différents temps de présentation avec un objet translatant devant le FoE de l'observateur se déplaçant avec un *heading* de 5 degrés. Le biais d'estimation du *heading* induit par l'objet en translation s'élève à -0.75 degré. Les graphiques du bas représentent également les activations des neurones de l'aire MSTd cette fois-ci avec l'observateur suivant un *heading* de 5.5 degrés et un objet se déplaçant vers l'observateur ayant pour FoE 11.5 degrés. Le biais alors induit par ce mouvement de l'objet vers l'observateur est de 2.7 degrés.

des objets en mouvement dans le champ visuel, est explicable par la mise en commun de la sélectivité au mouvement et des composantes du flux optique au sein des aires MT et MST et qu'un traitement différentiel du mouvement n'est pas nécessaire pour ces tâches d'estimation.

Pour conclure ce chapitre et résumer son contenu, le système visuel humain constitue une riche et complexe hiérarchie permettant le traitement

des données reçues par les rétines. Les informations lumineuses captées d'abord par l'œil passent à travers différents étages de neurones au sein de la rétine avant d'être transmises au CGL et finalement au cortex visuel, décrivant ainsi la voie rétino-géniculo-striée. Une fois arrivée au cortex visuel, l'information est traitée par les différentes couches présentes et se sépare en deux voies de traitement visuel. En s'intéressant ici à la voie dorsale, impliquée notamment dans la perception du mouvement, il est remarqué qu'un traitement local puis global de l'information de mouvement est opéré au sein des aires MT et MST pour permettre d'extraire du flux optique différentes informations pertinentes pour la locomotion.Ce flux optique défini comme les composantes visuelles issues du mouvement observé permet alors de s'orienter et de correctement estimer sa trajectoire à l'aide de différents mécanismes hérités du traitement du flux optique. Ces mécanismes de flux optique tels que le *heading* ou le *flow parsing* étudiés au travers de différentes études et modélisations biologiques révèlent leur intérêt et leur efficacité pour le suivi de trajectoire ou l'aide à la locomotion. La question de l'émergence de cette sélectivité à la direction de mouvement et aux composantes de flux optique au sein du cortex visuel peut alors se poser. Ainsi le développement des neurones et de leurs champs récepteurs à cette sélectivité à la direction de mouvement et son apprentissage, qui a notamment pu être mis en évidence par [Clemens et al., 2012] chez le furet, s'effectue dans les premiers jours d'ouverture des yeux grâce à l'expérience visuelle. Ainsi les champs récepteurs de neurones observés et responsables de la sélectivité à la direction du mouvement sont d'abord aléatoires et se précisent au fur et à mesure d'un apprentissage visuel non supervisé finissant par répondre à différentes composantes du flux optique tels que les mouvement de translation verticaux ou horizontaux. Cependant, les mécanismes responsables du développement de cette sélectivité et son émergence restent incertains.

Le chapitre suivant aura pour objectif de présenter comment ces différents mécanismes de traitement du flux optique observés chez le vivant (et notamment chez le primate) peuvent être reproduits au sein de systèmes artificiels à partir de caméras asynchrones et de réseau de neurones événementiels.

Chapitre 3

Systèmes bio-inspirés, de la captation au traitement de l'information visuelle

Dans le chapitre précédent, j'ai décrit le fonctionnement général du système visuel chez le primate et notamment les mécanismes neuronaux impliqués dans le traitement du flux optique lors de la locomotion. Dans ce chapitre seront décrits les systèmes inspirés par la biologie afin de capter l'information visuelle et de la traiter pour la navigation et la locomotion. Ces différents systèmes, qualifiés de bio-inspirés, reprennent le fonctionnement et les traitements de la voie rétino-géniculo-striée afin de s'approcher au mieux des traitements effectués chez le primate.

Il sera décrit en premier lieu les systèmes permettant la captation de l'information visuelle et leur association aux modèles permettant le traitement de cette information, ici les réseaux de neurones artificiels. Il sera ensuite décrit le modèle des réseaux de neurones à impulsion et leur fonctionnement, modèle impliquant des processus de traitement bio-inspirés pour finalement décrire des modèles computationnels de traitement du mouvement et du flux optique faisant écho aux modèles biologiques présentés dans le chapitre précédent.

3.1 Capter l'information visuelle

La perception visuelle du monde qui nous entoure s'effectue grâce à notre système visuel et implique notamment la rétine capable de capter les informations lumineuses comme détaillé dans le chapitre précédent. C'est sur ce même principe de captation de l'information lumineuse que reposent aujourd'hui les caméras modernes afin d'enregistrer et de restituer le mouvement d'une scène filmée. Ces types de capteurs visuels dits standards voient leur principe d'acquisition inchangé depuis l'invention du Kinématographe par Thomas Edison et William Dickson en 1888 et celle du Cinématographe par les frères Lumière en 1895. Ce principe repose sur l'acquisition d'images à intervalles de temps réguliers qui, regardées les unes à la suite des autres retransmettent la décomposition du mouvement observé, illustré par la figure 3.1. Cette captation standard effectuée par des capteurs aussi appelés *frame-based* s'est améliorée au fil des années à l'aide de l'évolution constante des technologies employées, et notamment du passage d'une captation analogique à numérique, passant de la traditionnelle pellicule aux capteurs photosensibles.

FIGURE 3.1 – Chronophotographies décomposant le mouvement d'un cheval au galop, par Eadweard Muybridge, Animal Locomotion, 1887.

Ces capteurs, constitués aujourd'hui d'une matrice CMOS pour leur grande majorité, mise au point par [Fossum, 1997] dans les années 1990,

permettent un traitement vidéo efficace et rapide pour de multiples applications (captation, analyse, communication, etc.). Bien que leur évolution permet une amélioration grandissante en termes de qualité, de résolution, ou encore de fréquence d'enregistrement, ces capteurs *frame-based* reposent sur toujours le même procédé : la capture d'une image à intervalles réguliers, procédé pouvant présenter des limites résultantes de leur nature synchrone. Ces limites peuvent être caractérisées par une redondance des données capturées pouvant amener à des besoins énergétiques et de mémoire élevés, des latences induites par le temps nécessaire entre chaque image capturée, entraînant une éventuelle perte d'information durant ces laps de temps, ou encore du flou de mouvement dû à une fréquence de capture des images non adaptée [Censi et al., 2015, Delbruck, 2016].

Palier aux limites que présente la caméra alors dite synchrone implique de s'affranchir de cette nature et de son fonctionnement reposant sur une fréquence d'horloge. Cela peut alors s'effectuer en se rapprochant d'un traitement déjà effectué par la rétine, exploitant les changements au sein d'une scène visuelle, consistant à capturer de manière asynchrone, c'est-à-dire événementielle, l'évolution de la scène.

La caméra événementielle

Une caméra événementielle, aussi appelée caméra asynchrone ou caméra neuromorphique est un capteur d'images réagissant aux changements de luminance de la scène qu'il est en train de filmer [Lichtsteiner et al., 2008, Posch et al., 2011, Brandli et al., 2014, Son et al., 2017]. Contrairement aux caméras classiques qui envoient un flux continu de données à une fréquence bien définie, ces caméras ont la particularité de ne renvoyer sous forme d'événements ou *spike*, que les éléments ayant changé entre deux instants temporels dans la scène. De ce fait, chaque pixel qui compose ces caméras fonctionne de manière asynchrone et indépendante signalant les changements au moment exact où ils apparaissent. Ces changements correspondent à des variations de luminance au sein de la scène observée, ainsi si un pixel voit sa valeur de luminance diminuer il émettra alors un évé-

nement dit 'OFF' témoignant d'un assombrissement local et, à l'inverse, un événement dit 'ON' témoignant d'un éclaircissement local, valeur déterminée selon la réponse logarithmique des pixels de ces caméras (voir la figure 3.2 et l'équation 3.1 où I représente l'intensité de la luminance d'un pixel aux coordonnées (x, y) et aux temps (t) et $(t + 1)$, et I_{thresh} le seuil de changement de l'intensité nécessaire à l'émission d'un spike). Cette méthode de traitement correspond à ce qui est effectué par la rétine qui traite les informations captées par les photorécepteurs sous la forme d'impulsions électriques qui sont transmises à ses différentes couches (cellules bipolaires et ganglionnaires) puis au nerf optique, rendant ainsi les caméras événementielles des systèmes bio-inspirés.

$$|log(I_{x,y,t+1}) - log(I_{x,y,t})| \geq I_{thresh} \tag{3.1}$$

Les caméras utilisant ce processus de détection de variation de luminance afin de ne pas encoder une image complète à une fréquence donnée mais bien de n'intégrer que les changements détectés sont catégorisées comme étant des Dynamic Vision Sensor (DVS). Les caméras DVS se reposent sur un traitement de l'information en plusieurs étapes : une première étape se chargeant de la captation de la luminosité de la scène filmée, une deuxième étape détectant les variations de luminance et une dernière pour la génération des spikes 'ON' ou 'OFF'. Cette génération de spikes selon le mouvement capturé est représentée dans la figure 3.2.

Plusieurs modèles de caméras événementielles existent aujourd'hui (Prophesee, iniVation, Insightness, Samsung, CelePixel) et suivent principalement un traitement basé sur la différence de luminance de chacun des pixels du capteur de la caméra dans le temps [Posch et al., 2014, Steffen et al., 2019, Gallego et al., 2022]. Bien qu'étant bio-inspirés par la rétine, ces modèles manquent un facteur important de cette inspiration du modèle visuel humain : les filtres spatiaux. En effet l'information visuelle une fois captée par les photorécepteurs de la rétine est filtrée une première fois en passant par les cellules ganglionnaires avant de l'être une seconde fois en passant par le corps géniculé latéral pour finir au cortex visuel (voir le chapitre pré-

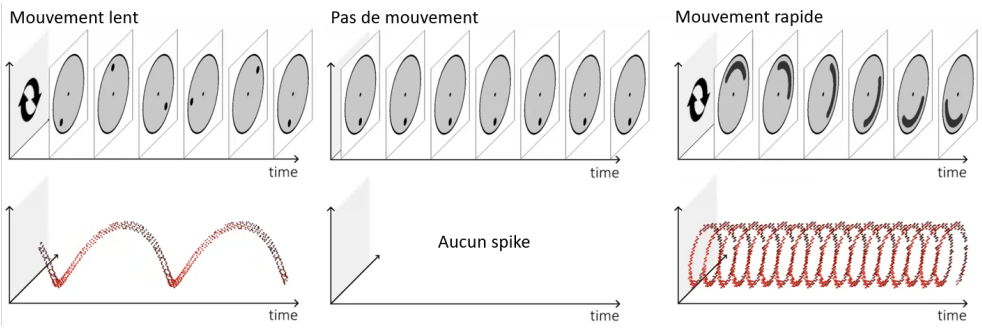


FIGURE 3.2 – Comparaison du fonctionnement d'une caméra *frame-based* et d'une caméra *event-based* ou événementielle établie par [Mueggler et al., 2014]. La scène est ici constituée d'un point noir sur un disque blanc qui effectue des rotations dans le sens horaire pendant trois conditions de mouvement et à travers deux caméras, une caméra synchrone au-dessus et une caméra asynchrone en-dessous. En premier lieu (à gauche), le disque décrit un mouvement de rotation lent : la caméra synchrone récupère la position du disque noir aux instants définis par son horloge pouvant alors rater des positions intermédiaires, là où la caméra asynchrone montre la position du disque noir à chaque instant, mettant en évidence la limite de latence temporelle entre les deux caméras. En deuxième lieu (au centre), le disque ne bouge plus : la caméra synchrone continue d'intégrer l'information visuelle dans sa totalité alors que la caméra asynchrone n'émet aucun spike puisqu'aucun mouvement n'est observé, mettant en évidence la redondance d'informations émise par la caméra synchrone par rapport à la caméra asynchrone. En troisième lieu (à droite), le disque décrit un mouvement similaire que dans le premier cas, mais plus rapide : la caméra synchrone fait alors apparaître un flou de mouvement là où la caméra synchrone reste capable d'intégrer la position du disque noir, mettant en évidence une meilleure plage dynamique.

cédent). Ces différents filtrages spatiaux prennent la forme de différences de Gaussiennes (DoG) [Olshausen and Field, 1997, Hubel and Wiesel, 1962]. Tout cela survient en conservant leurs propriétés rétinotopiques, c'est-à-dire que leurs propriétés spatiales restent inchangées : deux champs récepteurs proches l'un de l'autre sur la rétine le resteront ici aussi [Enroth-Cugell and Robson, 1966, Rodieck, 1965].

De plus, l'utilisation de tels filtres permet également la réduction du bruit favorisant ainsi une meilleure détection des mouvements [Petkov and

Subramanian, 2007]. Ce sont de telles propriétés que l'on retrouve dans la caméra proposée par *Yumain* : le *Spike Event Sensor*, décrite par [Debat, 2021, Debat et al., 2021] et retrouvable en annexe A, et qui est utilisée ou simulée pour la génération de certains de nos jeux de données.

L'utilisation de telles caméras s'inspirant alors du fonctionnement de la rétine nécessite un traitement adapté des données renvoyées. Cette bio-inspiration des caméras dans leur captation et leur perception de la scène filmée amène à également s'inspirer de la suite de la chaîne de traitement du système visuel humain. Le traitement des sorties des caméras peut alors venir alimenter un réseau de neurones artificiels.

3.2 Les réseaux de neurones artificiels

Reproduire le comportement et le fonctionnement du cerveau humain, c'est ce qui a motivé les réflexions autour de ce que l'on appelle les réseaux de neurones artificiels. Occupant aujourd'hui une place importante au sein des méthodes de l'intelligence artificielle, cette section s'intéresse à leur découverte et leurs applications entraînant des variations dans leur nature et leur architecture.

Principe et fonctionnement

Pour bien comprendre le principe et le fonctionnement des réseaux de neurones artificiels, il faut s'intéresser à leur origine, comment ils furent inventés et leur évolution pour en arriver à la technologie utilisée actuellement.

Les premiers réseaux de neurones artificiels ont été inventés en 1943 par Warren S. McCulloch et Walter Pitts, mathématiciens et neuroscientifiques. Dans [McCulloch and Pitts, 1943], ils décrivent comment un agencement de neurones artificiels, en s'inspirant des neurones biologiques, a pu permettre son organisation pour former à partir de cet arrangement un réseau où les neurones le composant sont capable d'échanger de l'information.

Le neurone biologique

Les neurones biologiques sont des entités excitables connectées les unes aux autres ayant pour rôle de transmettre les informations sous forme de signaux électriques. Chaque neurone est composé de dendrites, d'un corps cellulaire, d'un axone et de terminaisons neuronales (voir la figure 3.3). Les dendrites sont les portes d'entrée d'un neurone. C'est à cet endroit, au niveau des synapses, que le neurone reçoit les signaux provenant des neurones qui le précède faisant varier son potentiel de membrane. Ces signaux peuvent être de différentes nature : soit ils seront excitateurs, soit inhibiteurs. Lorsque le potentiel de membrane, variant en fonction de la somme de ces signaux, dépasse un certain seuil, le neurone s'active et produit alors un signal électrique à son tour, déchargeant son potentiel de membrane. Ce signal circule le long de l'axone pour être envoyé à son tour à d'autres neurones de notre système nerveux, neurones qui fonctionnent exactement de la même manière.

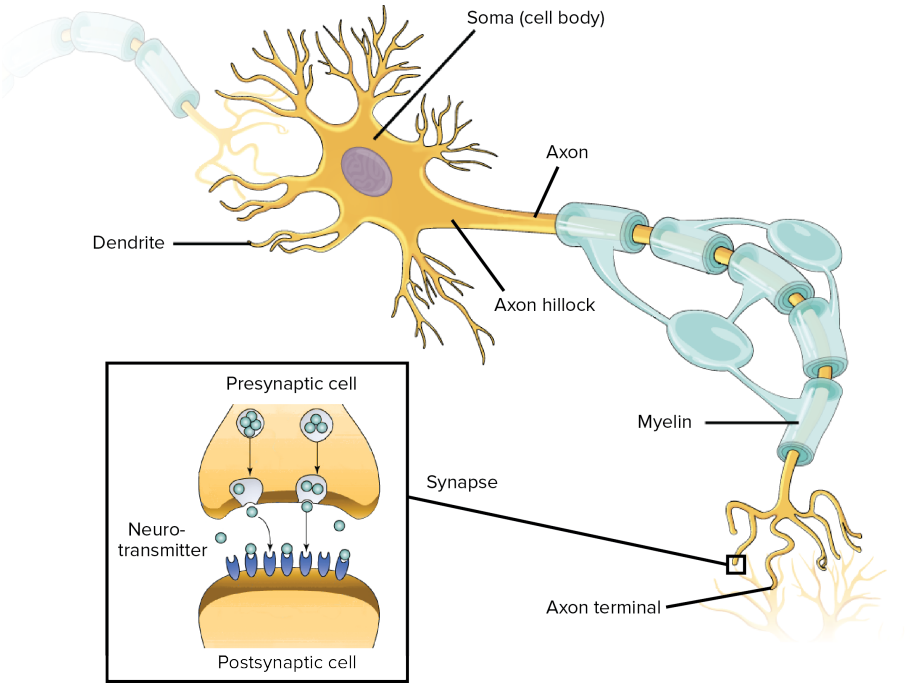


FIGURE 3.3 – Schéma d'un neurone biologique tiré de [Clark et al., 2018].

Le neurone artificiel

C'est en s'inspirant du modèle biologique que McCulloch et Pitts ont modélisé le principe et le fonctionnement d'un neurone artificiel. Ce fonctionnement est donné par une fonction de transfert qui reçoit des signaux en entrée et en retourne en sortie. A l'intérieur de cette fonction, on trouve deux grandes étapes.

La première est une étape d'agrégation : on somme toutes les entrées du neurone en les multipliant par un coefficient représentant le poids synaptique, positif si le signal est excitateur, et à l'inverse négatif si le signal est inhibiteur. Dans cette phase d'agrégation on obtient donc une fonction ayant pour expression l'équation 3.2, elle même illustrée par la figure 3.4.

$$f = \sum x_i w_i \tag{3.2}$$

La seconde étape est une étape d'activation : on s'intéresse au résultat du calcul effectué durant l'étape d'agrégation, et si celui-ci dépasse un certain seuil pré-définit, alors le neurone s'active et retourne une valeur de sortie, fonctionnement décrit par l'équation 3.3 retranscrivant le comportement de cette fonction d'activation, appelée fonction d'Heaviside.

$$\begin{cases} y = 1 & \text{si } f \geq 0 \\ y = 0 & \text{sinon} \end{cases} \tag{3.3}$$

Ces deux étapes ont ainsi permis de développer les premiers neurones artificiels, aujourd'hui appelés *Threshold Logic Unit* (TLU). Ce nom vient de leur conception à ne traiter que des entrées logiques et binaires. McCulloch et Pitts ont pu montrer qu'avec ce modèle il était possible de reproduire certaines opérations logiques telles que les fonctions 'AND' et 'OR'. Ils ont pu également démontrer qu'en connectant plusieurs de ces fonctions les unes aux autres, comme on peut l'observer chez le vivant, il est possible de résoudre n'importe quel problème de logique booléenne.

Si ici est décrit un des premiers modèle de neurone artificiel qu'est le TLU, bien d'autres ont pu voir le jour au fil des années. Néanmoins, conserver la vraisemblance biologique inspirée par le neurone biologique n'a pas

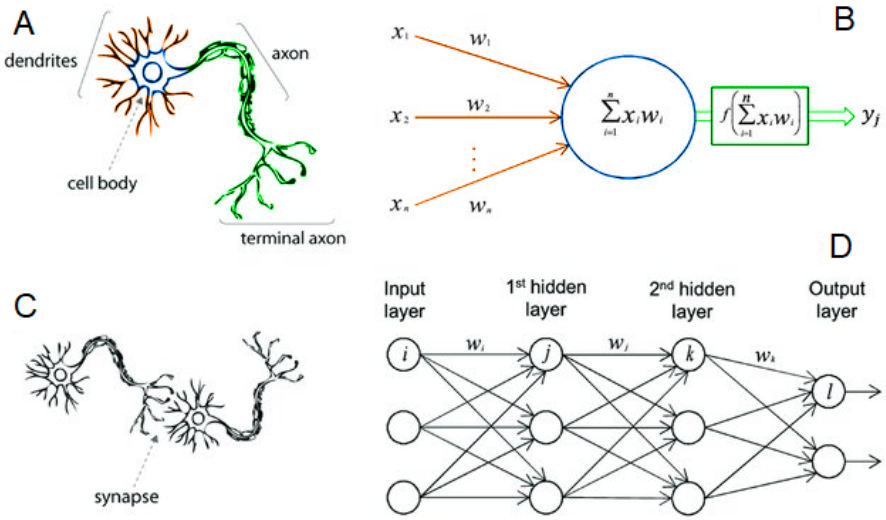


FIGURE 3.4 – Comparaison entre neurones biologiques et artificiels adaptée de [Meng et al., 2020]. A) Neurone biologique. B) Neurone artificiel. C) Réseau de neurones biologiques avec connexions synaptiques. D) Réseau de neurones artificiels.

toujours été une priorité, privilégiant une efficacité computationnelle à une nature bio-inspirée. Toutefois, ici seront présentés des modèles artificiels de neurones biologiques conservant leurs principales caractéristiques. La fonction d'un neurone artificiel inspiré par le modèle biologique reste d'intégrer un potentiel d'action reçu en entrée, faisant varier ses paramètres associés aux synapses le connectant aux autres, afin de décharger un potentiel d'action à son tour.

Le modèle Hodking-Huxley

Le modèle Holdking-Huxley (H&H) consiste en un ensemble d'équations différentielles non linéaires décrivant la relation entre le courant traversant le neurone par ses canaux ioniques et son potentiel de membrane [Hodgkin and Huxley, 1952b, Hodgkin and Huxley, 1952a, Hodgkin et al., 1952, Hodgkin and Huxley, 1952c]. Premier modèle décrivant un neurone par ses propriétés de capacitance et de conductance, il permet de décrire avec finesse certains comportement du neurone biologique comme l'augmentation du potentiel

de membrane en fonction du courant afférant induit par les spikes en entrée du neurone. Cependant sa complexité due à ses nombreux paramètres et équations (c.f. équations 3.4) le rend peu utilisable si l'on souhaite favoriser la rapidité d'exécution plutôt que la précision des calculs.

$$\begin{cases} C_m \dfrac{dV(t)}{dt} = -\sum\limits_i I_i(t, V) \\ I(t, V) = g(t, V) \cdot (V - V_{eq}) \\ g(t, V) = \bar{g} \cdot m(t, V)^p \cdot h(t, V)^q \end{cases} \tag{3.4}$$

Ici, la première équation décrit la relation entre le potentiel $V(t)$ et le courant I associé à la capacité membranaire C_m. La deuxième exprime le courant I en fonction du temps t et du potentiel de membrane V faisant ainsi apparaître la conductance g. Finalement la troisième équation exprime la conductance g en fonction de t et de V incluant les paramètres $\bar{g}$ la conductance maximale et les fractions d'activation et d'inactivation m et h déterminant le débit d'ions traversant les canaux de la membrane du neurone.

Le modèle *integrate-and-fire* (IF)

Introduit par Louis Lapicque en 1907 et décrit par [Abbott, 1999], ce modèle constitue l'un des premiers modèles de neurone, représenté par son potentiel de membrane V évoluant en fonction de son courant I à travers le temps, décrit par l'équation 3.5.

$$I(t) = C\frac{dV(t)}{dt} \tag{3.5}$$

Ainsi, lorsque le neurone décrit par ce modèle reçoit un courant, le potentiel de membrane augmente jusqu'à atteindre son potentiel de seuil V_{th}. Une fois atteint, le neurone émet un spike et son potentiel de membrane se voit revenir à son état de repos. Ainsi sa fréquence d'émission de spikes est directement contrôlée par le courant et peut être ajustée par l'ajout d'une période réfractaire t_{ref} empêchant le neurone d'émettre pendant cette période. La fréquence d'émission est alors décrite par l'équation 3.6.

$$f(I) = \frac{I}{CV_{th} + t_{ref} I} \tag{3.6}$$

Ce modèle décrit le modèle *integrate-and-fire* parfait et sera par la suite amélioré afin de se rapprocher du modèle biologique tout en gardant son aspect simplistique et facile d'intégration, n'étant décrit que par une équation différentielle linéaire de premier ordre. Il donnera alors naissance aux modèles de neurones *adaptative integrate-and-fire*, *exponential integrate-and-fire*, ou encore *leaky integrate-and-fire*, modèle qui sera décrit en détail puisque proposé au sein du SNN présenté plus loin.

Différents types de réseaux de neurones artificiels

Suite à la publication de [McCulloch and Pitts, 1943], il y eut un engouement pour l'intelligence artificielle. Cependant même si leur modèle pose les bases de ce qu'aujourd'hui sont les réseaux de neurones artificiels, il ne dispose pas, par exemple, d'algorithmes d'apprentissage pour trouver la valeurs des coefficients synaptiques. C'est en 1957 que le domaine se voit prendre un nouveau tournant par l'invention du Perceptron par [Rosenblatt, 1958].

Le Perceptron

Le modèle du Perceptron ressemble de très près à celui établi par McCulloch et Pitts décrit précédemment. Il s'agit d'un neurone artificiel qui s'active lorsque la somme pondérée de ses entrées dépasse un certain seuil. Le Perceptron dispose également d'un algorithme d'apprentissage lui permettant de trouver les valeurs de ses paramètres synaptiques afin d'obtenir les sorties qui conviennent. Pour développer cet algorithme, Frank Rosenblatt s'est inspiré de la théorie de Hebb. Cette théorie suggère que lorsque deux neurones biologiques sont excités conjointement ils renforcent alors leur lien synaptique [Hebb, 1949]. C'est ce que l'on appelle la plasticité synaptique. A partir de cette idée - développée plus loin dans ce manuscrit - Frank Rosenblatt a proposé un algorithme d'apprentissage qui consiste à

entraîner un neurone artificiel sur des données de référence pour que celui-ci renforce ses paramètres synaptiques à chaque fois qu'une entrée est activée en même temps que la sortie présente dans ces données. La formule décrite dans l'équation 3.7 permet la mise à jour de ces paramètres synaptiques en calculant la différence entre la sortie de référence et la sortie produite par le neurone, et en multipliant cette différence par la valeur de chaque entrée et un facteur d'apprentissage positif.

$$W = W + \alpha(y_{ref} - y)X \tag{3.7}$$

Dans l'équation 3.7 se retrouve W les paramètres d'apprentissage synaptique, X les valeurs d'entrée du neurone, y la sortie produite par le neurone, y_{ref} la sortie de référence attendue en sortie du neurone, et α la vitesse d'apprentissage. Cette équation est à mettre en relation avec la figure 3.5 où l'on retrouve ces mêmes paramètres afin de décrire ce modèle linéaire.

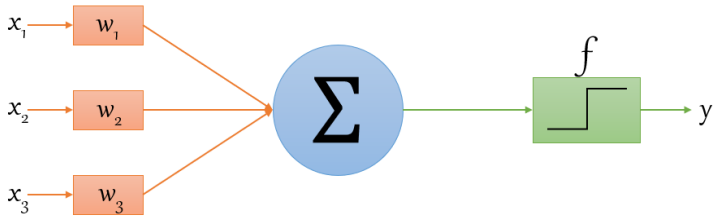


FIGURE 3.5 – Le modèle Perceptron. Ici les données en entrée x_n sont multipliées avec leur poids synaptique associé w_n avant d'être sommées. Cette somme passe à travers la fonction d'activation f pour en obtenir la sortie définie ici comme fonction d'Heaviside. Les poids sont finalement ajustés en fonction de la sortie obtenue selon l'équation 3.7.

Le Perceptron Multicouche

La nature du modèle Perceptron (illustré figure 3.7, décrit comme un modèle linéaire - c'est-à-dire ne pouvant catégoriser des données non linéairement séparables - peut différencier les fonctions logiques 'AND' et 'OR',

linéairement séparables, mais se retrouve dans l'impossibilité de classifier des points répartis selon la fonction logique 'XOR', comme montré figure 3.6.

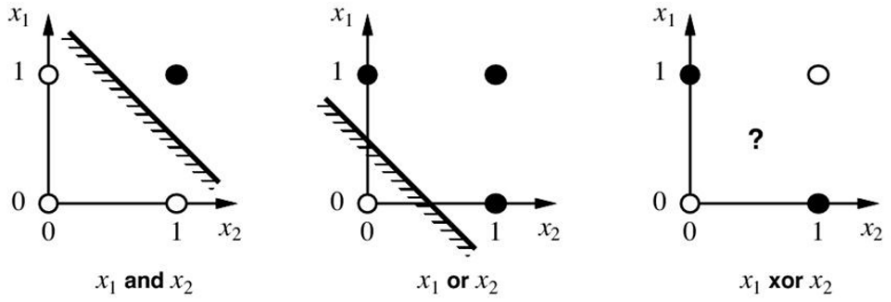


FIGURE 3.6 – Séparabilité linéaire des modèles de fonctions logiques 'AND', 'OR' et 'XOR'. La limite du modèle linéaire est mise en évidence par l'impossibilité de séparer les points du modèle 'XOR' à l'aide d'une simple droite, adapté de [Chandradevan, 2017]

Afin de résoudre ce problème non linéaire, [Rumelhart et al., 1986] proposent de rajouter une couche supplémentaire de neurones Perceptron afin de ne plus avoir une couche d'entrée et une couche de sortie, mais d'y intercaler une ou plusieurs couches, dites couches cachées, définissant ainsi un réseau de neurones artificiels. De plus, afin d'ajuster les poids synaptiques dans ce modèle, ils proposent la rétropropagation de l'erreur de la couche de sortie à travers les couches précédentes jusqu'à la couche d'entrée.

Cette solution de rétropropagation consiste donc à déterminer comment la sortie du réseau varie en fonction des paramètres présents dans chaque couche du modèle. Pour cela, on calcule une chaîne de gradients indiquant comment la sortie varie en fonction de la dernière couche, puis comment la dernière couche varie en fonction de l'avant-dernière, jusqu'à arriver à la couche d'entrée du réseau. Grâce à ces gradients on peut alors mettre à jour les paramètres de chaque couche de telle sorte à ce qu'ils minimisent l'erreur de la sortie du modèle et la réponse attendue en utilisant la formule de descente de gradient décrite par l'équation 3.8.

$$W = W - \alpha \frac{\partial Erreur}{\partial W} \tag{3.8}$$

L'algorithme de descente de gradient est un algorithme d'optimisation permettant de trouver la valeur minimum de toute fonction dite convexe, en se rapprochant progressivement de celle-ci. Ce type d'algorithme, largement utilisé dans les domaines de l'apprentissage supervisé, permet alors de minimiser la fonction de coût, c'est-à-dire l'erreur calculée entre la sortie obtenue et la sortie attendue. En effet, trouver le minimum de la fonction de coût décrivant l'erreur d'un réseau revient à corriger l'estimation de sortie et à trouver le meilleur modèle convenant au réseau auquel elle est associée.

La descente de gradient va ainsi permettre d'ajuster les paramètres synaptiques du réseau en prenant en compte l'erreur donnée par la fonction de coût et le taux d'apprentissage associé à chacun des paramètres synaptiques que l'on retrouve dans l'équation 3.8 avec W les différentes valeurs des paramètres synaptiques, α le taux d'apprentissage, et $Erreur$ la valeur de l'erreur renvoyée par la fonction de coût.

Développer et entraîner les réseaux de neurones artificiels les plus utilisés aujourd'hui peut alors se résumer de la sorte :

— Propagation avant : les données d'entrées reçues par la première couche sont propagées vers l'avant à travers toutes les couches du réseau afin de produire une sortie ;
— Fonction de coût : on calcule l'erreur entre la sortie obtenue et la sortie de référence que l'on désire retrouver ;
— Rétropropagation : on mesure comment la fonction de coût varie par rapport à toutes les couches du réseau de neurones en partant de la dernière jusqu'à la première ;
— Descente de gradient : on corrige chaque paramètres du modèle grâce à l'algorithme de descente de gradient avant de revenir à la première étape pour recommencer un cycle d'entraînement.

Au fil du temps, le modèle du Perceptron multicouches a continué d'évoluer notamment avec l'apparition de nouvelles fonctions d'activation telles

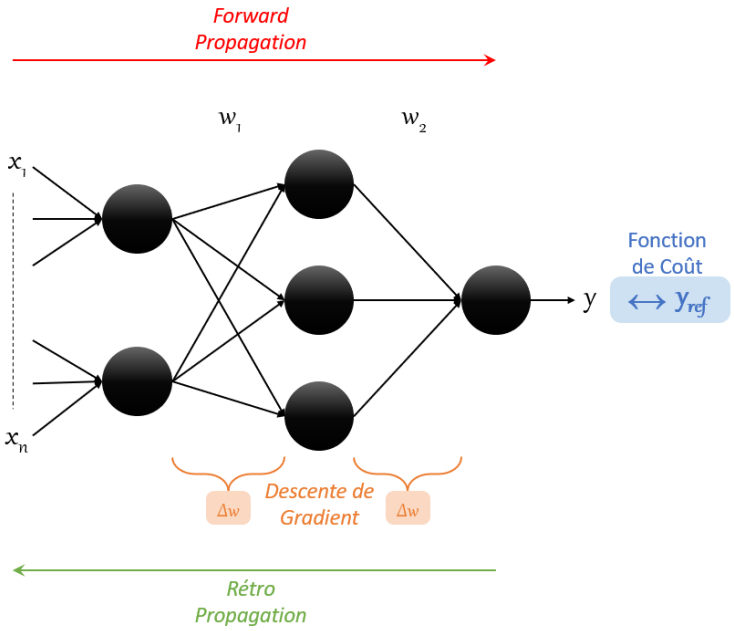


FIGURE 3.7 – Le modèle du Perceptron multicouche illustrant les différentes étapes les plus utilisées pour l'entraînement des réseaux de neurones artificiels. Ici x représente les différentes entrées du réseau et w les différentes connexions synaptiques. Par propagation avant (ou *forward propagation*) les entrées passent à travers le réseau passant par les neurones et les synapses afin d'obtenir une sortie y. Cette sortie est comparée à la sortie attendue y_{ref} afin de minimiser la fonction de coût. Elle est ensuite rétropropagée à travers le réseau et les valeurs des connexions synaptiques sont corrigées par descente de gradient par un facteur Δw.

que la fonction Logistique, la fonction Tangente Hyperbolique (Tanh), ou encore la fonction Rectified Linear Unit (ReLU). Ces fonctions ont aujourd'hui remplacé la fonction Heaviside, fonction d'activation du Perceptron, car elles offrent de meilleures performances. Au cours des années 1990, les premières variantes des réseaux multicouches voient le jour. [LeCun and Bengio, 1995] présentent les premiers réseaux de neurones convolutifs, réseaux capables de reconnaître et de traîter des images en introduisant des filtres dits de *convolution* et de *pooling*.

D'autres types de réseaux de neurones artificiels vont aussi voir le jour comme les réseaux de neurones récurrents, toujours une variante du Percep-

tron multicouches, permettant le traitement des problèmes de séries temporelles comme la lecture de texte [Mikolov and Zweig, 2012] ou la reconnaissance vocale [Graves et al., 2013].

Depuis, les réseaux de neurones artificiels se sont de plus en plus complexifiés et sont devenus de plus en plus efficaces, capables de résoudre des tâches et des analyses de plus en plus complexes. Néanmoins, cette recherche de la meilleure performance a fait s'éloigner les réseaux de neurones artificiels de leur première inspiration, le cerveau humain. En effet les mécanismes utilisés dans les réseaux convolutifs ou récurrents par exemple se basent moins sur les mécanismes que l'on peut retrouver dans le cerveau, faisant appel à des fonctions d'activation ou d'apprentissage, ou des modèles de neurones s'éloignant des modèles biologiques. C'est pour retrouver des notions plus biologiques qu'ont été développés les réseaux de neurones à impulsions, ou réseaux de neurones événementiels.

Le réseau de neurones à impulsions

Afin de pouvoir traiter et intégrer les impulsions transmises par les caméras événementielles, l'utilisation d'un réseau de neurones artificiels est privilégiée puisque traitant l'information spikante de l'entrée jusqu'à sa sortie. Toujours dans cette idée de système bio-inspiré et pour se rapprocher au plus près des processus biologiques effectués par le système visuel du primate, les réseaux de neurones à impulsions (SNN : *Spiking Neural Network*) se montrent être un choix pertinent : à l'instar des neurones naturels, les neurones composant un SNN se basent sur l'intégration des impulsions reçues, telles que les événements envoyés par une caméra événementielle, et l'émission de nouveaux événements selon leur modèle et leur règle d'apprentissage. Ainsi, est proposé un SNN à apprentissage non supervisé suivant la règle d'apprentissage *Spike Timing-Dependent Plasticity* (STDP), et composé de neurones intégrateurs à fuite ou *Leaky Integrate-and-Fire* (LIF) reprenant et s'inspirant de travaux tels que [Bichler et al., 2012, Diehl and Cook, 2015, Masquelier and Thorpe, 2007, Paredes-Vallés et al., 2020, Falez, 2019].

Le modèle de neurone LIF

Le modèle de neurone proposé est basé sur le modèle *Leaky, Integrate and Fire* (LIF) [Abbott, 1999, Gerstner and Kistler, 2002]. Un neurone LIF possède un potentiel de membrane V_m, un potentiel de repos V_{rest}, une résistivité membranaire R_m, une constante de temps τ_m, et un courant I. Quand ce type de neurone reçoit une impulsion en entrée, son potentiel de membrane augmente. Si le neurone n'est pas déjà à son potentiel de repos et n'intègre aucun événement, son potentiel de membrane chute progressivement et de manière exponentielle. Un neurone LIF présente aussi un potentiel de seuil V_thresh. Supposons que le potentiel de membrane dépasse le potentiel de seuil. Dans ce cas, le neurone émettra un événement, et son potentiel de membrane restera ensuite à son potentiel de repos durant une certaine période appelée période réfractaire. L'équation 3.9 et la figure 3.8 décrivent ce modèle.

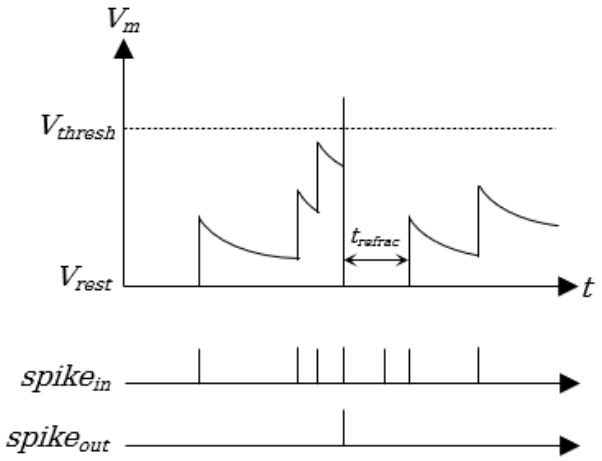


FIGURE 3.8 – Le modèle du neurone LIF. Le potentiel de membrane V_m varie en fonction des spikes entrant $spike_{in}$. Quand la valeur de seuil V_{thresh} est atteinte, le neurone LIF émet un spike $spike_{out}$ et revient à son potentiel de repos V_{rest}. Le potentiel du neurone LIF reste à cet état de repos durant une période réfractaire t_{refrac}.

$$\tau_m \frac{d}{dt} V_m(t) = -(V_m(t) - V_{rest}) + R_m I(t) \tag{3.9}$$

La règle d'apprentissage

La règle d'apprentissage utilisée dans notre SNN est une règle de *Spike Timing-Dependent Plasticity* (STDP) additive. La STDP, originellement découverte par [Bi and Poo, 1998, Markram et al., 1997, Sjöström et al., 2001], est censée avoir un rôle significatif dans les mécanismes d'apprentissage du cerveau selon [Dan and Poo, 2004]. Plus largement, la règle d'apprentissage STDP repose sur le mécanisme de neuroplasticité. En effet, les synapses font partie des responsables de la plasticité cérébrale et de l'apprentissage grâce à leur mécanisme de remodelage et de reconfiguration [Raisman, 1969]. Cette plasticité synaptique décrivant une modification de la force des connexions synaptiques entre les neurones se voit être décrite par Donald Hebb et résumée de la façon suivante : *"cells that fire together, wire together"*, autrement dit "des neurones s'excitant ensemble se lient entre eux" [Hebb, 1949]. Ainsi la plasticité synaptique se voit être définie par un changement des paramètres d'entrée et de sortie d'une synapse entre deux neurones décrivant ainsi deux composantes dites présynaptique pour les paramètres d'entrée de la synapse correspondant au neurone duquel provient la connexion, et postsynaptique pour les paramètres de sortie de la synapse correspondant au neurone recevant les informations venant de la connexion.

Plusieurs études se sont alors penchées sur la modélisation du phénomène de plasticité synaptique dans un effort de se rapprocher du traitement neuronal biologique faisant émerger deux catégories : les modèles fréquentiels et les modèles événementiels.

Modèle fréquentiel

Le modèle fréquentiel pour la modélisation du phénomène de plasticité synaptique est le plus courant [Dayan and Abbott, 2001]. Ce modèle définit le signe et l'amplitude de la plasticité synaptique en fonction de la fréquence de décharge des neurones présynaptiques et postsynaptiques. Il se présente suivant l'équation 3.10 où W_s représente le poids synaptique de la synapse s, x_s la fréquence d'émission du neurone présynaptique, et y la

fréquence d'émission du neurone postsynaptique. Tous les modèles fréquentiels dépendent alors de cette formule et peuvent y apporter des termes en plus comme y inclure le taux d'apprentissage [Linsker, 1986], un terme de décroissance [Oja, 1982], ou encore un seuil d'activation modifiable pour le modèle BCM [Bienenstock et al., 1982].

$$\frac{dW_s}{dt} = f(x_s, y, W_i) \tag{3.10}$$

Modèle événementiel

Le modèle événementiel repose sur la règle d'apprentissage STDP. Son apprentissage est basé sur le temps écoulé entre 2 événements venant de 2 neurones connectés l'un à l'autre par une synapse. Un événement d'entrée provenant du premier neurone est émis juste avant un événement de sortie venant du second neurone provoquant un mécanisme de Long-Term Potentiation (LTP), ayant pour effet de renforcer le poids de la connexion synaptique qui leur est associée. De la même manière, supposons qu'un événement d'entrée venant du premier neurone soit émis juste après un événement de sortie venant du second neurone. Dans ce cas, le poids de la connexion synaptique reliant les deux sera diminué, induisant un mécanisme de Long-Term Depression (LTD). Ce changement de poids synaptique est exprimé ci-dessous par l'équation 3.11 proposée par [Masquelier and Thorpe, 2007] et illustré par la figure 3.9.

$$\Delta w = \begin{cases} -A_{LTD} + W \cdot e^{\frac{\Delta t}{\tau_{LTD}}} & \Delta t \leq 0 \\ A_{LTP} + W \cdot e^{\frac{-\Delta t}{\tau_{LTP}}} & \Delta t > 0 \end{cases} \tag{3.11}$$

Ici, est retrouvé Δw le changement de poids synaptique, A l'amplitude d'apprentissage, W le poids actuel, τ la fenêtre temporelle d'apprentissage, et Δt le temps entre un événement d'entrée et de sortie.

Ce modèle événementiel mis en avant par la règle STDP ne permet alors que la prise en compte de paires de spikes, correspondant au dernier spike émis par le neurone présynaptique et spike émis par le neurone postsynaptique. D'autres modèles événementiels existent donc tout en suivant cette

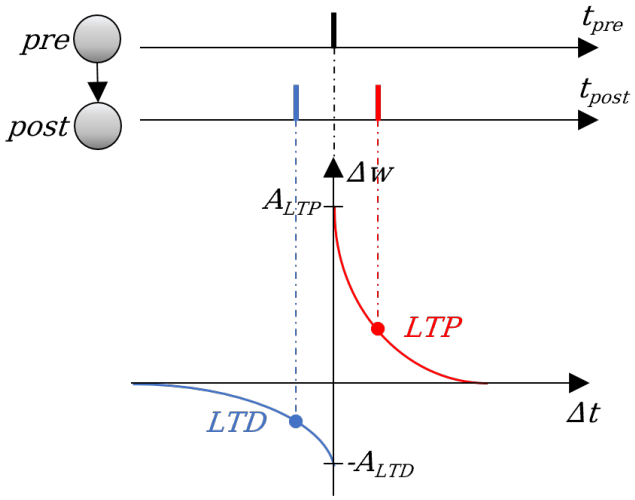


FIGURE 3.9 – Illustration de la règle d'apprentissage STDP. Lorsqu'un neurone pré-synaptique spike juste avant un neurone post-synaptique, leur poids synaptique associé est augmenté par un facteur Δw par LTP. Plus l'intervalle $\Delta t = t_{post} - t_{pre}$ entre ces deux spikes est faible tout en restant positif, plus cet augmentation sera importante (courbe rouge ici). A l'inverse, leur poids synaptique est diminué par LTD lorsque le neurone pré-synaptique émet un spike après le neurone post-synaptique (courbe bleue ici) et traduit par $\Delta t < 0$.

même règle d'apprentissage comme la triplet-STDP par [Pfister and Gerstner, 2006] qui prend en compte des triplets de spikes au lieu de paires, ou la *Mirrored* STDP par [Burbank, 2015] proposant des règles STDP temporellement opposées selon le type de connexion synaptique.

La règle d'apprentissage STDP étant une règle d'apprentissage non supervisée, des alternatives à ce modèle sont proposées sous la forme de règles semi-supervisées et supervisées. C'est dans un soucis de performances des SNNs que s'inscrit par exemple la Reward-modulated STDP ou R-STDP par [Mozafari et al., 2019] comme étant semi-supervisée. Ainsi les décisions correctes prises par le SNN sont dirigées vers la règle STDP et à l'inverse vers une règle anti-STDP. Aussi la rétropropagation se voit être adaptée aux SNNs, notamment par l'architecture S4NN par [Kheradpisheh and Masquelier, 2020] et l'utilisation du *surrogate gradient* [Neftci et al., 2019, Pellegrini et al., 2020, Woźniak et al., 2020] mais s'éloignant de la plausibilité biolo-

gique, où la STDP reste candidat à l'observation chez le vivant [Zhang et al., 1998, Crochet and Petersen, 2006, Jacob et al., 2007, Young et al., 2007, Feldman, 2012].

Finalement, la mise en relation entre les mécanismes de modèle de neurone LIF et de règle d'apprentissage STDP permet alors la création du SNN qui se retrouve proposé dans les études présentées par la suite. Ce SNN est donc doté de neurones artificiels recevant de l'information sous forme d'impulsions (ou spikes) venant agir sur le potentiel membranaire de ceux-ci. Ces neurones, répartis en groupes, forment alors des couches, responsables de l'intégration à plusieurs niveaux des données d'entrée. Entre ces couches et reliant les neurones entre eux se retrouvent les connexions synaptiques, auxquelles est appliquée la règle d'apprentissage non-supervisée STDP, régissant alors le poids synaptique des différentes connexions et venant s'ajuster en fonction des entrées et sorties des neurones LIF. Selon les données en entrée, les connexions synaptiques peuvent être arrangées de différentes façons : soit de manière complètement connectée où tous les neurones d'une première couche sont reliés à ceux de la seconde, soit de manière rétinotopique (ou convolutionnelle) où les neurones entre deux couches du réseau se voient n'être connectés qu'à certaines parties permettant d'obtenir des champs récepteurs locaux de l'activité événementielle transmise en entrée du réseau.

Pour s'assurer que les différents neurones d'une même population, faisant partie de la même couche dans l'architecture du réseau, n'apprennent pas les mêmes propriétés spatio-temporelles, un mécanisme d'inhibition latérale a été implémenté (voir [Chauhan et al., 2018]). À chaque fois qu'un neurone émet une impulsion, il empêche tous les autres neurones de la même couche que lui d'émettre à leur tour, induisant alors une période réfractaire forcée.

Cette architecture peut alors être comparée aux différents modèles de réseaux de neurones artificiels utilisant les mêmes mécanismes de traitement de l'information et d'apprentissage.

3.3 Réseaux de neurones à impulsions pour le traitement du flux optique

Ces dernières années ont vu l'augmentation du nombre d'études utilisant des données événementielles pour la vision par ordinateur, avec des performances parfois bien meilleures que celles obtenues par des approches plus classiques à partir de caméras *frame-based* pour des applications telles que la reconnaissance d'objets [Neil and Liu, 2016, Stromatias et al., 2017] ou l'odométrie visuelle [Gallego and Scaramuzza, 2017, Nguyen et al., 2019]. Ces études sont toutes basées sur des réseaux convolutionnels profonds ou des SNNs, couplés à des approches de classification ou d'apprentissage supervisé [Lakshmi et al., 2019]. Par exemple, [Zhu et al., 2019] utilise un réseau de neurones artificiels (ANN) pour prédire le flux optique à partir de données événementielles acquises par une caméra montée sur une voiture se déplaçant dans un environnement urbain (voir aussi [Zhu et al., 2018b]). Dans [Lee et al., 2020], les auteurs se servent du même jeu de données mais traité à l'aide d'un réseau hybride ANN/SNN produisant de meilleurs résultats pour l'estimation du flux optique à l'aide d'un apprentissage supervisé grâce à l'utilisation d'un mécanisme de type integrate-and-fire pour les neurones de leur réseau. Le SNN est ainsi capable de déterminer le flux optique en tous points de la scène observée pour chaque événement enregistré (voir figure 3.10).

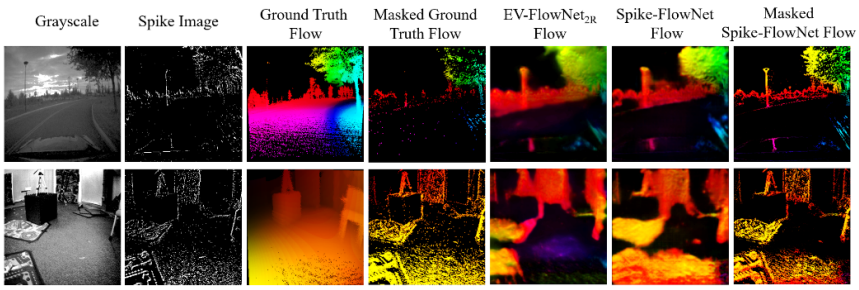


FIGURE 3.10 – Evaluation du flux optique obtenu par [Lee et al., 2020] comparé au jeu de données utilisé ([Zhu et al., 2018a]) et à [Zhu et al., 2019].

Des approches alternatives basées sur l'apprentissage non-supervisé ont été également développées. Dans [Bichler et al., 2012], les auteurs démontrent que la sélectivité au mouvement peut être apprise par des SNNs héritant d'une règle d'apprentissage STDP, bio-inspirée et non-supervisée. Leur réseau est capable de discriminer la direction de mouvement sur des données événementielles mais aussi de compter les véhicules roulant dans différentes voies d'une autoroute à partir de données collectées à l'aide d'un DVS. Leurs résultats permettent alors de montrer qu'une simple règle d'apprentissage biologique non supervisée telle que la STDP permet de générer une sélectivité à des patterns complexes de spikes à une échelle globale comme le montre leur architecture figure 3.11 mais aussi à une échelle locale par une configuration topologique du réseau dite rétinotopique.

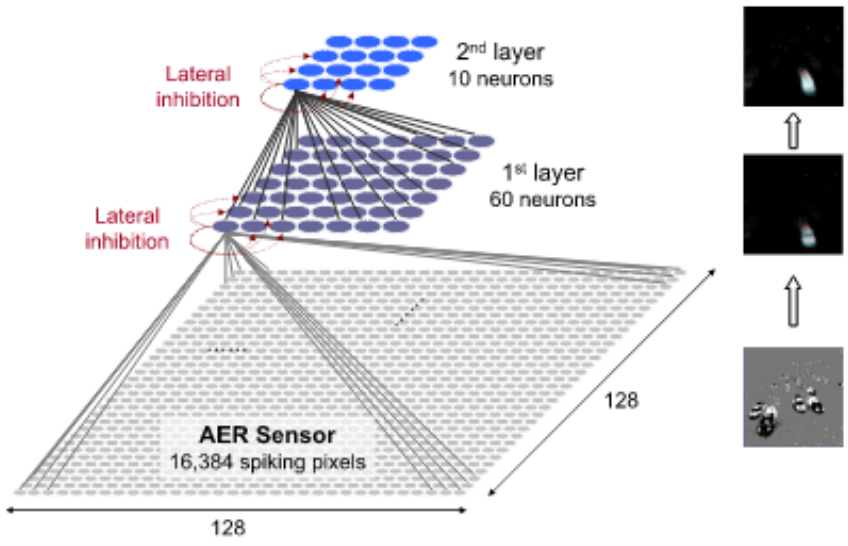


FIGURE 3.11 – Architecture du SNN proposé par [Bichler et al., 2012] pour l'apprentissage de patterns spatio-temporels complexes à une échelle globale. Il est constitué de deux couches connectées de manière *feed-forward* avec une inhibition latérale entre les neurones d'une même couche. Sur la droite est représenté un champ récepteur précisant les événements favorisant les réponses d'un des neurones du réseau.

Dans [Paredes-Vallés et al., 2020], un réseau hiérarchique profond basé sur de nombreuses couches et incluant des délais de transmissions est ca-

pable d'estimer les patterns de mouvement d'objets se déplaçant après un apprentissage non-supervisé par STDP. Ce réseau reste cependant complexe et présente différentes approches de traitement et formatage de données à travers de multiples couches et neurones mais permet néanmoins une sélectivité à la position et à la vitesse des objets grâce à l'apprentissage et au traitement du flux optique estimé au travers des filtres spatio-temporels décrits par les champs récepteurs des neurones après apprentissage.

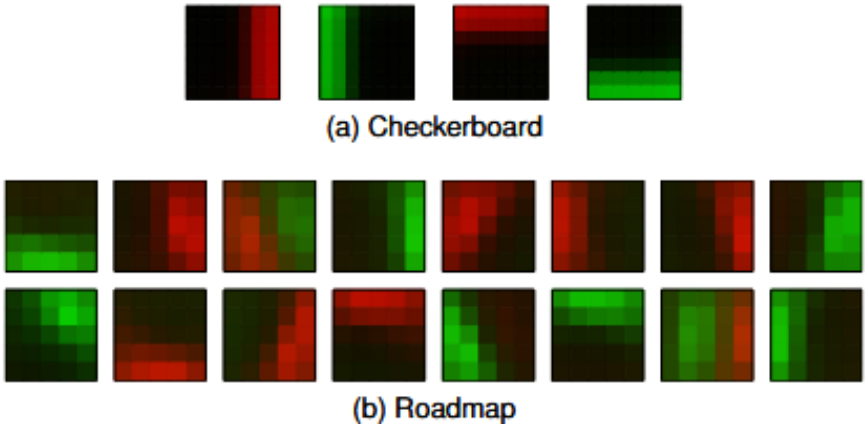


FIGURE 3.12 – Différents champs récepteurs de neurones obtenus après apprentissage sur un jeu de données simulées représentant un damier (a) et sur des séquences d'un jeu de données réelles (b). La couleur verte représente la sélectivité aux événements 'ON' et le rouge la sélectivité aux événements 'OFF'. Ainsi sur le premier jeu de données les neurones répondent à des mouvements simples, ceux-ci se complexifient pour faire apparaître des mouvements combinant des événements 'ON' et 'OFF'.

[Barbier et al., 2021] font également intervenir des délais de transmissions synaptiques au sein d'un SNN présentant moins de couches et de paramètres que celui présenté par [Paredes-Vallés et al., 2020]. Ainsi les auteurs présentent un réseau capable d'apprendre les orientations et le mouvement à partir d'une règle d'apprentissage STDP et de neurones LIF. Ce réseau s'inspire de la biologie en faisant intervenir des cellules simples et complexes pour les différentes couches qui le composent. Son entraînement est effectué dans diverses conditions, d'abord avec des simulations de barres, puis des formes filmées à l'aide d'une caméra événementielle, et finalement dans

un contexte de navigation avec une entrée stéréoscopique à l'aide de deux caméras événementielles appairées sur une plateforme mobile. Les champs récepteurs des neurones après entraînement montrent alors une sélectivité aux disparités locales de mouvement et de direction (voir figure 3.13).

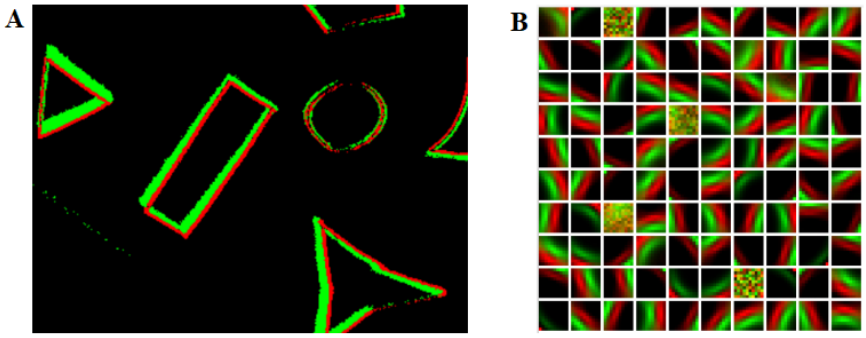


FIGURE 3.13 – A) Événements générés par la caméra événementielle utilisée pour la capture de formes. B) Champs récepteurs obtenus après entraînement sur la vidéo représentée présentant les différentes formes, d'après [Barbier et al., 2021].

Tout aussi récemment, [Debat et al., 2021] s'intéressent à la détection de balles en mouvement et à prédire leurs points d'arrivées à partir d'une captation des trajectoires avec une caméra événementielle et un apprentissage de celles-ci à l'aide d'un SNN. Par l'utilisation de délais pour une discrimination des vitesses et une architecture multicouche convolutionnelle permettant une sélectivité aux propriétés spatiales, l'étude montre l'efficacité d'un réseau tel que leur SNN dôté d'une règle STDP simplifiée à prédire où une balle atterrira après un lancer. Les champs récepteurs des neurones servent alors de filtres pour la détection du mouvement et se spécifient à certaines zones de la scène filmée lors de passes de balles comme le montre la figure 3.14.

C'est à partir de ces différentes études que le développement d'une nouvelle architecture de SNN s'est basé, qui, quand associée à une règle STDP, apprend à extraire les propriétés du flux optique, notamment le mouvement de soi durant la navigation à partir de données événementielles collectées dans des conditions de locomotion naturelle. Il sera alors question dans le

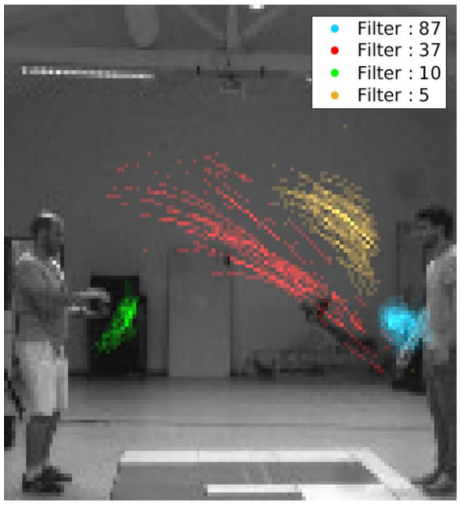


FIGURE 3.14 – Le champ visuel de la caméra filmant une scène où une balle est passée d'une personne à l'autre et dont les événements sont extraits afin de servir d'entrée au SNN. Les couleurs en superposition désignent les zones spatiales auxquelles certains neurones sont devenus sélectifs après un entraînement sur plusieurs passes. Il est remarqué que ces zones correspondent aux bras des lanceurs d'une part, et la balle dans ses trajectoires ascendante et descendante d'autre part, des deux côtés de la scène où les deux lanceurs se trouvent. Ici ne sont représentés que l'activité à laquelle réagissent quatre filtres, tiré de [Debat et al., 2021].

chapitre suivant d'un SNN plus simple que ceux proposés dans ces précédentes études et ainsi plus simple à configurer et optimiser, permettant la mise en évidence d'un apprentissage des composantes du flux optique et l'émergence d'une sélectivité au mouvement.

Chapitre 4

Extraction événementielle des indices de navigation par apprentissage non-supervisé des composantes du flux optique

Ce chapitre met en application les éléments présentés dans le chapitre précédent. Je décris ici ma première étude sur le développement et l'utilisation d'un réseau de neurones impulsionnels pour l'apprentissage des composantes du flux optique. Les méthodes et résultats présentés ont fait l'objet de communications scientifiques aux conférences *Bernstein Computational Neuroscience 2021* et *VISAPP 2022*, en qualité de poster et d'article respectivement [Fricker et al., 2021, Fricker et al., 2022].

4.1 Rationnel de l'étude

Durant la locomotion, différents schémas de flux optique projetés sur la rétine sont utilisés par de nombreuses espèces animales afin de contrôler leurs directions et vitesses de déplacement. Comme nous l'avons vu dans le chapitre 2, chez les primates, le flux optique est traité par un réseau hiérarchique qui, dans un premier temps extrait les composantes locales du

mouvement, puis les combine afin de déterminer les propriétés globales du mouvement, afin de permettre l'extraction des paramètres de navigation. Ce traitement est très efficace en termes de consommation d'énergie, l'information étant transmise par le système visuel sous forme de spikes, et il est généralement admis que le cerveau humain ne requiert que 20 watts pour son fonctionnement [Mink et al., 1981]. Reproduire ces mécanismes neuronaux au sein de systèmes artificiels et embarqués pourrait avoir de sérieuses implications dans les domaines industriels (e.g., pour le développement de véhicules autonomes) et cliniques (e.g., pour l'aide à la navigation chez les personnes aveugles). Ces dernières années ont vu l'émergence de nombreuses études où le flux optique a pu être traité avec une perception bio-inspirée grâce au développement de caméras événementielles, similaires par leur fonctionnement à la rétine humaine. La transmission des données acquises par ces caméras est asynchrone et présente une forte résolution temporelle (jusqu'à la milliseconde [Posch et al., 2014]), ce qui induit un potentiel très avantageux pour de l'application temps réel.

Une façon naturelle de traiter les spikes émis par ces caméras événementielles est d'utiliser les SNNs comme vu dans le chapitre précédemment à travers différents modèles décrits. Ces réseaux favorisent une faible consommation d'énergie et peuvent être implémentés sur des puces dites neuromorphiques telles que l'Intel Loihi [Davies et al., 2018] ou l'IBM TrueNorth [Akopyan et al., 2015]. Apprendre avec des SNNs peut être fait sans supervision. Dans le premier cas, la nature discrète des spikes rend difficile toute estimation des paramètres du réseau par rétropropagation, bien que de récentes avancées telle que la méthode du *surrogate gradient descent* mènent à des résultats prometteurs [Neftci et al., 2019, Zenke et al., 2021]. L'apprentissage nécessite souvent une grande quantité de données labellisées, et généraliser cette labellisation à travers plusieurs contextes visuels n'est pas toujours garanti. Une approche par apprentissage non-supervisé permet une alternative aux méthodes supervisées de par leur flexibilité aux modifications d'entrée, et ne nécessite pas une labellisation des données d'entrée.

Ici sera décrit un réseau de neurones à spikes efficace et fonctionnel apprenant à extraire les composantes utiles du flux optique pendant la

navigation par l'intermédiaire d'une règle d'apprentissage bio-inspirée et non-supervisée. Il sera montré qu'après entraînement, les sorties du réseau deviennent sélectives aux différentes composantes du flux optique, notamment les composantes translationnelles, rotationnelles et radiales. Finalement, l'activation de ces différentes sorties pourra être utilisée pour la prédiction du mouvement de soi au cours de tâches de navigation.

4.2 Méthode

Le processus de traitement présenté ici consiste en un SNN intégrant des données événementielles en adéquation avec celles reçues par la rétine au cours de la navigation. Il sera d'abord décrit les différents types de données événementielles utilisées en entrée du réseau, puis les propriétés de ce dernier ainsi que la règle d'apprentissage utilisée pour l'entraînement. Finalement, seront détaillées les méthodes d'évaluations employées pour caractériser les propriétés du réseau après apprentissage.

Jeux de données

Deux différents jeux de données événementielles ont été utilisés pour l'apprentissage du réseau.

Simulations simplifiées de patterns de flux optique

Afin de caractériser la capacité du réseau à apprendre les composantes du flux optique, des simulations simplifiées ont été mises au point, consistant en quatre disques blancs sur fond noir se déplaçant dans différentes directions. Chacun des quatre disques prend place au sein d'un des quadrants du champ visuel en effectuant des mouvements de translation (vers la gauche, la droite, le bas, ou le haut), de rotation (dans le sens horaire ou anti-horaire), ou d'expansion/contraction.

Chaque disque a un diamètre de six pixels pour une taille totale du champ visuel de 32×32 pixels, laissant une taille de quadrant de 16×16 pixels. Chaque simulation a été générée à partir de 16 instants temporels

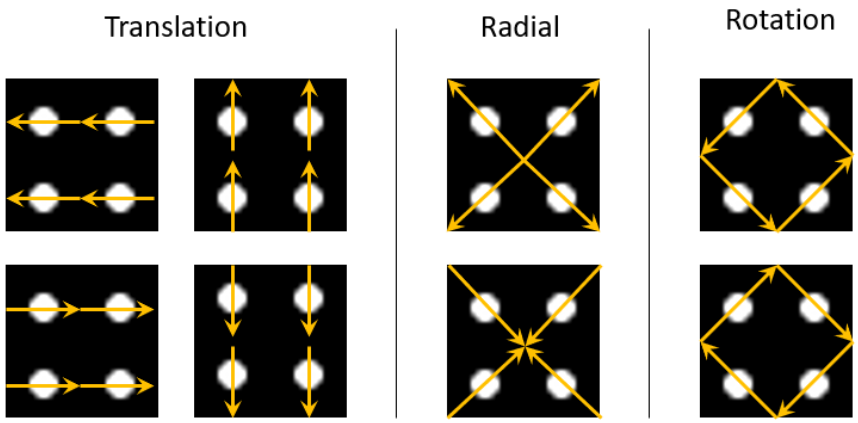


FIGURE 4.1 – Les différentes simulations des composantes du flux optique. Ici sont représentées les composantes translationnelles, radiales et rotationnelles par le mouvement de quatre disques blancs sur fond noir chacun dans un quadrant de l'image. Les flèches jaunes décrivent le mouvement des disques au cours du temps.

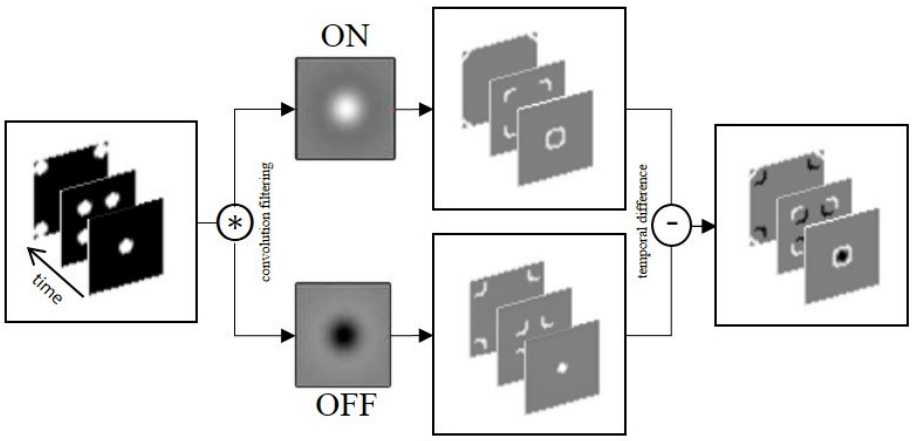


FIGURE 4.2 – Génération et pré-traitement des simulations de flux optique. Les différentes images composant les simulations passent par un filtrage spatial par filtres DoG ('ON' et 'OFF') et une différence temporelle pour la génération des spikes correspondant aux différentes composantes.

présentés successivement en utilisant différentes vitesses : 120, 240, ou 480 pixels par seconde, induisant ainsi des séquences vidéos de respectivement 133 ms, 67 ms, ou 33 ms. Au total, 800 simulations ont été utilisées pour

l'entraînement du SNN (100 pour chaque condition de mouvement du flux optique), présentées dans un ordre aléatoire. Ces séquences sont filtrées par des noyaux spatiaux du type différence de Gaussienne (DoG) de taille 5×5 pixels. Les événements dits spikes sont générés à chaque fois que la valeur de sortie obtenue après différence entre deux instants temporels consécutifs dépassait une valeur de seuil donnée. Pour chaque spike, l'information transmise au SNN contient sa signature temporelle, spatiale, et sa polarité.

Ces simulations constituent un excellent premier cadre afin d'évaluer cette approche car elles permettent le contrôle total des patterns de flux optique transmis au SNN. Elles permettent aussi de caractériser la robustesse du modèle au bruit en manipulant le ratio signal sur bruit (SNR) dans les spikes d'entrée. Cette modification du SNR s'effectue en ajoutant des spikes aléatoirement en entrée.

Données événementielles collectées durant la navigation contextuelle

Le second jeu de données utilisé a été acquis à partir d'une caméra événementielle placée sur la tête d'un participant humain se déplaçant au sein d'un environnement urbain. La caméra utilisée est une DAVIS qui se caractérise par une résolution spatiale de 240×180 pixels avec une latence minimum de 3 μs et une plage dynamique de 130 dB. Pour entraîner et évaluer le réseau, les analyses se concentrent sur la partie centrale du champ visuel capté par la caméra, un carré de 60 pixels de côté. Cette restriction permet de limiter le nombre de spikes en entrée et ainsi d'améliorer la vitesse de traitement.

Une unité de mesure inertielle (IMU) a été utilisée pendant l'acquisition des données. Elle donne accès aux valeurs d'accélération et de vitesse angulaire sur les trois axes de trajectoire du piéton, acquises à une fréquence de 1 kHz. La trajectoire suivie durant l'acquisition des données consistait en une large boucle que la participant parcourait en marchant vers l'avant en empruntant des virages vers la droite et vers la gauche. La durée totale du parcours était de 133 secondes. Aussi, le chemin emprunté présentait plus

de virages vers la droite que vers la gauche, ayant pour conséquence une sur-représentations des mouvements vers l'avant et vers la droite au sein du jeu de données.

Format des données

Que les données proviennent des simulations ou des caméras événementielles, les spikes sont encodés selon l'*Adress-Event Representation* (AER). Ce format contient alors les coordonnées spatiales du spike, son marquage temporel ainsi que sa polarité. Ils sont par la suite regroupés par paquets de même durée et transmis au SNN via un ordonnanceur traitant tous les spikes entrants. Après le traitement de chaque paquet de spikes, le réseau entre dans une période de repos. Durant celle-ci, les potentiels de membrane de tous les neurones sont remis à leur valeur de repos.

Architecture du réseau de neurones

Si la section sur les réseaux de neurones à impulsions du chapitre 3 s'attarde sur les mécanismes que l'on retrouve au sein de ce type de réseau, ici est décrit comment ceux-ci interagissent entre eux afin de permettre un apprentissage non-supervisé. Le réseau alors présenté, mis au point et utilisé pour cette étude est un SNN composé de deux couches de neurones incluant un mécanisme d'inhibition latérale. Cette structure minimaliste, de par son nombre réduit de couches, permet de limiter le nombre de paramètres utilisés et peut ainsi faciliter l'éventuelle implémentation du système sur une puce neuromorphique et de limiter sa consommation d'énergie. La première couche est organisée de façon rétinotopique : chacun des neurones la composant reçoit des spikes ne venant que d'un quadrant du champ visuel. Cette couche comporte 64 neurones de type LIF (16 pour chaque quadrant) recevant leurs spikes au format AER via l'ordonnanceur. Les neurones composant la seconde couche, au nombre de 64, reçoivent des spikes de l'ensemble des neurones de la première couche (cette couche est dite '*fully connected*'). Les connexions synaptiques entre les neurones des deux couches sont modifiées en suivant la règle d'apprentissage non supervi-

sée de STDP. Afin d'empêcher les neurones d'apprendre les mêmes motifs, un mécanisme d'inhibition latérale s'ajoute entre les neurones d'une même couche. Ainsi lorsqu'un neurone émet un spike, il empêche tous les autres neurones de la même couche d'émettre à leur tour jusqu'à la fin de la présentation de l'entrée ou du paquet de spikes en cours. La figure 4.3 schématise cette architecture.

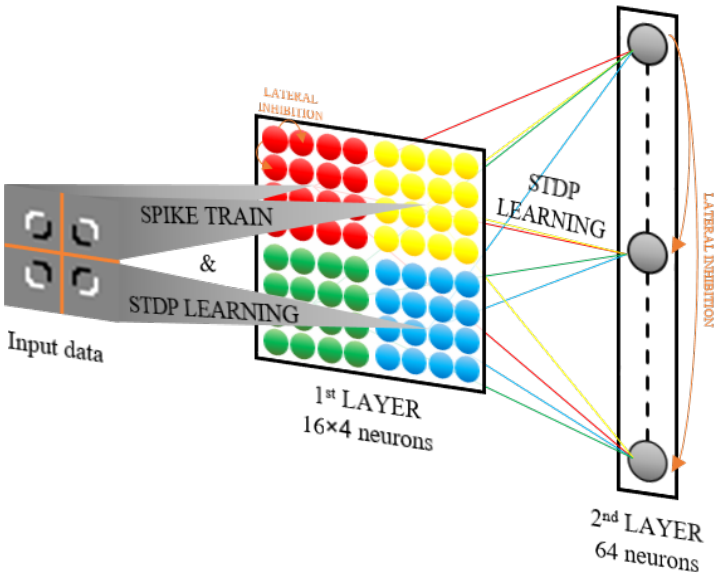


FIGURE 4.3 – Architecture du SNN utilisé. Les neurones de la première couche sont organisés de manière rétinotopique et ne reçoivent leurs spikes que de l'un des quatre quadrants du champ visuel (représentés ici par les différentes couleurs) traduisant une connexion *fully connected* par quadrant. Les neurones de la seconde couche sont *fully connected* aux sorties de la première couche. L'apprentissage non supervisé est d'abord présent entre l'entrée et la première couche. Après convergence des poids synaptiques, les spikes sont transmis à la seconde couche où la règle d'apprentissage est de nouveau appliquée.

Évaluation des performances

Pour caractériser la capacité de ce réseau à traiter et à reconnaître le flux optique, différentes mesures d'évaluation sont utilisées. Tout d'abord

la caractérisation de la sélectivité du réseau après apprentissage : puisque les patterns de flux optique sont nombreux dans les vidéos d'entrée, les neurones devraient progressivement apprendre des réponses spécifiques à ces patterns. Cette sélectivité peut être caractérisée à partir des champs récepteurs et des réponses des neurones après apprentissage. Les champs récepteurs permettent la représentation graphique des poids synaptiques des synapses connectées en amont des neurones. Ainsi, observer un champ récepteur d'un neurone permet de visualiser ce qu'il a appris et ce pour quoi il réagit. Pour compléter ces observations, des matrices de confusion peuvent également être calculées à partir des résultats obtenus, en utilisant l'approche proposée par [Diehl and Cook, 2015]. Après apprentissage, les réponses des neurones de la seconde couche sont définis selon les différents patterns de flux optique. En coupant l'apprentissage et en présentant au réseau les différents patterns de flux optique connus, la réponse d'un neurone à un certain pattern indique alors sa sélectivité à celui-ci. Ainsi pour un neurone particulier, sa composante de flux optique préférée correspond à celle ayant induit le nombre de spikes le plus élevé à sa sortie. Le neurone va ainsi se voir attribuer le label correspondant à cette composante particulière. Une fois cette labellisation effectuée, prédire la composante de flux optique présentée revient à sélectionner le label associé aux neurones répondant le plus fréquemment au sein de la population. La matrice de confusion vient alors spécifier la distribution des labels associés aux différents pattern de flux optique attendus. Dans l'idéal, la matrice de confusion est la matrice identité.

4.3 Résultats

Je présente ici les résultats que j'ai obtenus avec les deux jeux de données événementielles décrits ci-dessus.

Premier jeu de données

Après apprentissage du jeu de données événementielles simulées de flux optique, 50 pourcent des neurones de la seconde couche du réseau de neurones à spikes ont développé une sélectivité aux composantes du flux optique. La figure 4.5 illustre les réponses de huit de ces neurones avant et après entraînement non supervisé par STDP. Ainsi les champs récepteurs (voir figure 4.4) sont tout d'abord aléatoires puis deviennent progressivement des structures distinctes et réactives aux différents patterns de flux optique après apprentissage. Par exemple, le premier neurone à gauche de la figure 4.4 est sélectif à des translations vers la droite. Les différentes colonnes, associées à leur échelle temporelle, présentent l'activité spikante des différents neurones en réponse aux différents patterns de flux optique, apparaissant en entête de ces colonnes. Avant apprentissage, chaque neurone se voit répondre à différentes conditions de flux optique, sans distinction. A l'inverse, après apprentissage, après avoir présenté 800 conditions différentes (100 pour chaque pattern de flux optique dans un ordre aléatoire), les réponses se font plus rares et plus distinctes, chaque neurone ne répondant qu'à une seule composante.

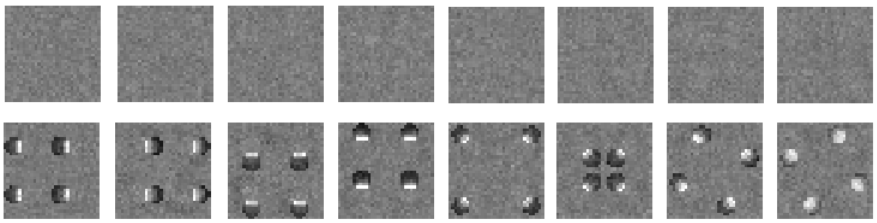


FIGURE 4.4 – Évolution des champs récepteurs des neurones de la seconde couche du SNN avant (en haut) et après apprentissage (en bas).

Ensuite, grâce aux matrices de confusion obtenues, les propriétés du réseau sont passées en revue. La figure 4.6 montre les différentes matrices obtenues sous différentes conditions de bruit (A) et un nombre variable de séquences présentées au réseau (B). En l'absence de bruit, la nature des patterns de flux optique est entièrement correctement classifiée grâce à l'ac-

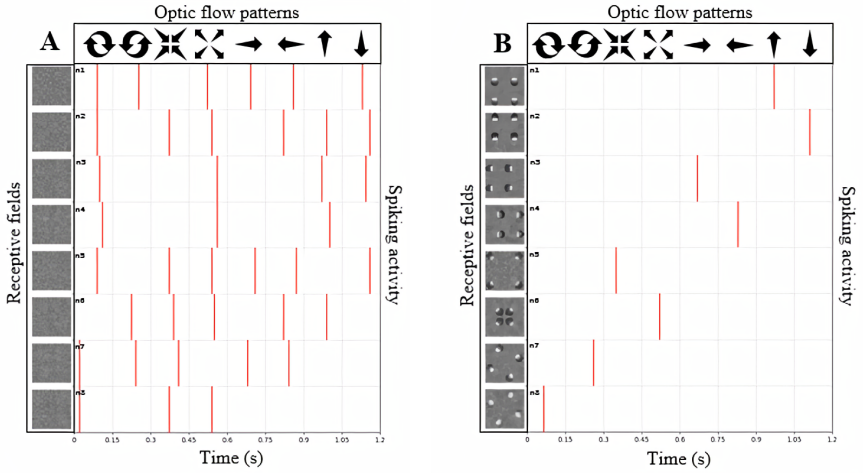


FIGURE 4.5 – L'activité neuronale du SNN avant (A) et après (B) apprentissage non supervisé du jeu de données événementielles simulées présentant différentes composantes du flux optique. Ici sont représentés les champs récepteurs des neurones dans la colonne la plus à gauche. Les régions blanches et noires correspondent respectivement aux changements de luminance positifs et négatifs. Les réponses aux différentes composantes du flux optique (représentées dans la ligne du haut) sont données par les spikes en rouge. Ces spikes sont associés par colonne à leur composante.

tivité spikante résultant de la phase d'apprentissage. Lorsque du bruit est ajouté, les performances diminuent mais reste néanmoins bien au-dessus du seuil de chance (12,5 pourcent), y compris pour les conditions testées les plus bruitées. Par exemple, 73 pourcent des prédictions sont correctes pour un rapport signal sur bruit de 0 dB. De plus, afin de s'assurer que l'apprentissage de ce réseau n'est pas basé sur la position initiale des disques mais bien sur leur mouvement, des simulations additionnelles ont été effectuées en rendant aléatoire les positions de départ des disques sur leurs trajectoires à la fois pendant la phase d'apprentissage et pendant la phase de test. Les performances de classification du réseau restent inchangées dans ce cas. Notamment, les composantes du flux optique sont toujours correctement estimées en l'absence de bruit, démontrant ainsi la capacité du réseau à apprendre les déplacements des disques.

Avec ce jeu de données simulées, les neurones de la seconde couche

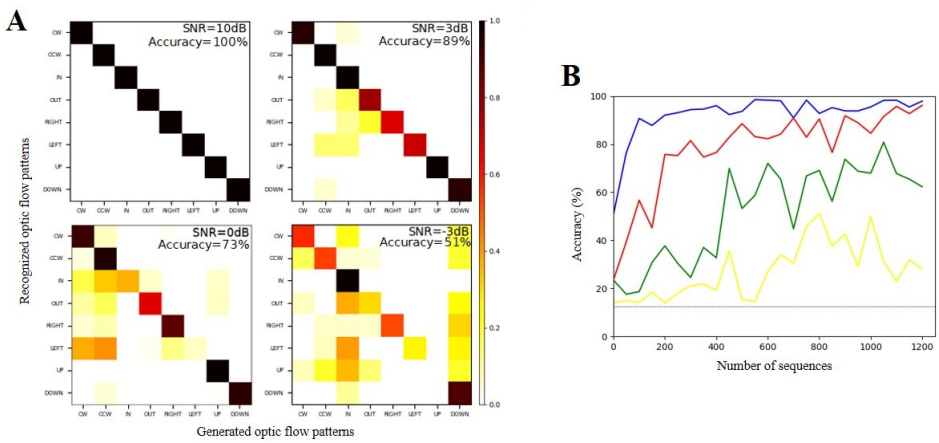


FIGURE 4.6 – Les performances observées du SNN sur le jeu de données événementielles simulées. (A) Les matrices de confusion obtenues après identification des différentes composantes du flux optique et sous quatre conditions de bruit différentes (SNR = 10, 3, 0, -3 dB). Les performances globales sont données aux coins supérieurs droits de chaque matrice. (B) Le niveau de précision du SNN (en pourcentages) en fonction du nombre de séquences présentées (allant de 1 à 2000) pour différentes valeurs du SNR (10 dB en bleu, 3 dB en rouge, 0 dB en vert, -3 dB en jaune). La ligne pointillée représente le niveau de chance (12,5 pourcent dans ce cas).

n'ayant pas convergé gardent leurs champs récepteurs à leur état d'initialisation aléatoire, même en augmentant le nombre de composantes présentées dans le jeu d'apprentissage. Cela peut être expliqué par le fait que les simulations ne présentent que huit conditions différentes et, dans ce cas, seulement un nombre limité de neurones est nécessaire afin d'extraire la direction de mouvement détectable dans les entrées. En outre, l'inhibition latérale présente dans le réseau bloque l'apprentissage de mêmes patterns au sein d'une même population de neurones. Cette même architecture est maintenant mise en œuvre pour l'apprentissage de données de navigation réelles.

FIGURE 4.7 – Le jeu de données de locomotion urbaine utilisé et capturé par [Mueggler et al., 2017]. La scène a été capturée simultanément par une caméra classique et une caméra événementielle. La ligne du haut montre la scène capturée par la caméra classique tandis que celle du bas montre les événements générés par la capture de la scène à l'aide d'une caméra événementielle. Chaque colonne représente le même instant temporel pour les deux caméras.

Second jeu de données

Ici, les performances du réseau sont testées à l'aide d'un jeu de données événementielles capturé lors de la locomotion par [Mueggler et al., 2017] et dont quelques captures sont montrées par la figure 4.7. En utilisant les données de référence fournies par l'IMU que sont les vitesses angulaires, le jeu de données peut être segmenté en trois catégories distinctes : les mouvements induits vers la gauche, vers la droite, et vers l'avant. Après apprentissage, tous les neurones de la seconde couche du réseau ont développé des réponses spécifiques au flux optique. Dans la figure 4.9, les activations de huit neurones sont montrées ainsi que leurs champs récepteurs avant et après l'apprentissage non supervisé par STDP. Les labels des composantes de flux optique sont représentés à l'aide de flèches en entête des figures. De la même façon qu'avec le jeu de données simulées, l'activité neuronale observée est d'abord aléatoire (les champs récepteurs des neurones sont bruités et répondent à toutes les conditions de flux optique). Après apprentissage,

les structures des champs récepteurs sont beaucoup plus détaillées et spécifiques et ne réagissent qu'à certaines catégories de flux optique (voir figure 4.8). Par exemple, le neurone représenté à la première ligne de la figure 4.9 est devenu sélectif au mouvement vers la gauche. Tous les neurones ne répondent qu'à une seule condition.

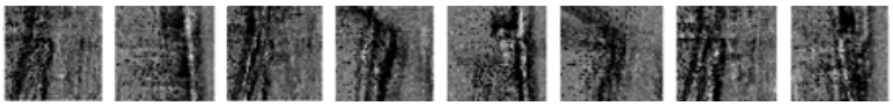


FIGURE 4.8 – Champs récepteurs remarquables des neurones du SNN après apprentissage sur des sections du jeu de données de [Mueggler et al., 2017] présentant des composantes translationnelles et radiales de flux optique.

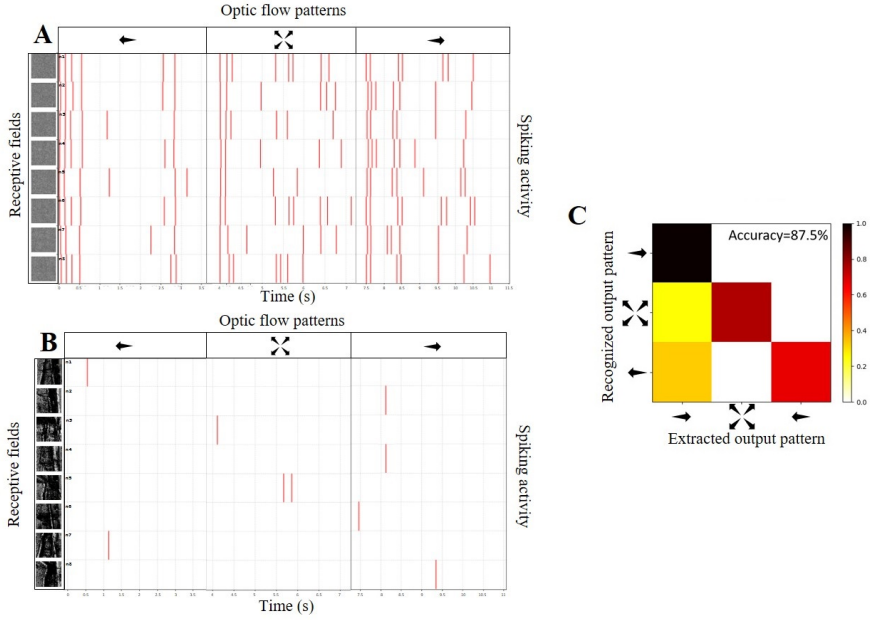


FIGURE 4.9 – Les performances observées du SNN sur le jeu de données événementielles de navigation. Les champs récepteurs (colonne de gauche) et l'activité neuronale (colonne de droite) sont montrés avant (A) et après (B) apprentissage de huit neurones représentatifs de la seconde couche du SNN. (C) La matrice de confusion après apprentissage. Le réseau est ici capable de déterminer avec 87,5 pourcent de précision les composantes de flux optique présentes dans les séquences capturées.

Nous testons ensuite la capacité du réseau à prédire les différentes conditions de flux optique à partir des spikes générés en sortie du réseau de neurones. Le processus est le même que celui décrit précédemment, après apprentissage les mêmes données sont représentées au réseau, cette fois-ci connues, et les réponses spikantes des neurones sont comptées. Chaque neurones se voit alors être attribué d'une classe correspondant au mouvement le faisant émettre le plus grand nombre de spikes avant de lui présenter l'intégralité du jeu de données pour la détection des trois composantes apprises. Dans ce cas, le niveau de chance est de 33,33 pourcent. La figure 4.9-C montre les performances du réseau sur ce jeu de données de navigation. L'apprentissage aboutit à un score moyen de 87,5 pourcent de classification correcte entre les trois différentes composantes de flux optique. Ce score est bien au-dessus du niveau de la chance, bien que certains mouvements vers la gauche (et moins fréquemment vers l'avant) soient interprétés comme étant des mouvements vers la droite. Comme mentionné plus tôt, les données de navigation utilisées contiennent essentiellement des déplacements vers la droite et vers l'avant, ce qui pourrait expliquer le biais de classification observé ici.

4.4 Conclusion

Nous avons présenté ici un réseau de neurones à spikes simple mais capable d'extraire les composantes du flux optique au sein de données événementielles. L'apprentissage dans ce réseau est complètement non supervisé et dépend d'une règle d'apprentissage bio inspirée, la STDP. Après convergence, les neurones présents dans le réseau deviennent sélectifs aux différentes composantes du flux optique, et leur activité au niveau de la population peut être utilisée afin de déterminer le mouvement de soi durant la navigation. Ces propriétés sont observables à la fois grâce aux données événementielles simulées, et aux données événementielles réelles collectées par une caméra DVS durant la locomotion.

Ce SNN comprend en tout 128 neurones dans ces deux couches tandis que les modèles de réseaux de neurones à impulsions précédemment décrits

dans le chapitre 2 [Paredes-Vallés et al., 2020], [Bichler et al., 2012], et [Diehl and Cook, 2015], utilisent respectivement 177, 266, et 6400 neurones dans leurs réseaux. Après apprentissage, le réseau présenté ici n'a besoin qu'entre un et dix spikes en sortie pour aboutir à une classification correcte d'un pattern de flux optique. Comme un spike à lui tout seul a une consommation estimée entre 700 et 900 pJ sur une puce neuromorphique [Indiveri et al., 2006, Aamir et al., 2018, Asghar et al., 2021], ce réseau aurait alors besoin au maximum de 7 à 9 nJ pour caractériser le mouvement de soi. De par sa simple architecture (seulement deux couches et 128 neurones) et sa faible consommation d'énergie estimée pour une implantation sur puce, ce réseau est donc un bon candidat pour des applications embarquées nécessitant un traitement du flux optique, par exemple, pour les véhicules autonomes ou pour la navigation avec assistance pour les patients aveugles.

Cependant, le jeu de données événementielles de navigation utilisé s'avère être complexe, et le chemin parcouru par le piéton portant la caméra événementielle très hétérogène. Les propriétés du réseau doivent alors être caractérisées par un jeu de données présentant des mouvements et composantes répartis de manière plus équilibrée présentant une complexité se situant entre les deux jeux de données présentés et utilisés ici. Dans le chapitre suivant sera alors décrit une différente méthode d'entraînement du réseau de neurones par l'élaboration d'un nouveau jeu de données utilisant à la fois simulation et données réelles.

Chapitre 5

Détection du mouvement au sein d'environnements numérisés en 3D à partir de réseau de neurones impulsionnels

Ce chapitre reprend le réseau de neurones à impulsions utilisé dans le chapitre précédent appliqué à un jeu de données qui diffère de ceux déjà proposés. Le jeu de données utilisé ici a été créé pour cette étude et est constitué de scènes capturées en 3D dans lesquelles se déplace une caméra événementielle restituant sa capture des spikes en fonction du déplacement voulu. Les données capturées sont ainsi apprises par le réseau pour la détection des composantes du flux optique au sein d'environnements contrôlés.

5.1 Méthode

L'utilisation du SNN proposé lors du chapitre précédent montre sa capacité à pouvoir apprendre les différentes composantes du flux optique lors de tâches présentant des complexités et situations différentes. Ainsi lors de l'apprentissage avec le jeu de données collecté par [Mueggler et al., 2017] présentant un contexte de navigation urbaine, celui-ci s'est avéré peu contrôlé

et trop hétérogène quant aux composantes présentes. Aussi l'évaluation des performances du SNN après apprentissage s'est révélée complexe en l'absence de données réelles pleinement exploitables. Il est alors présenté ici la réalisation d'un jeu de données permettant d'obtenir un espace contrôlé pour la génération de données événementielles lors de tâches de locomotion à travers différents environnements. Ces environnements sont capturés en 3D et numérisés dans un logiciel dédié permettant leur exploration à l'aide d'une caméra simulée dont les mouvements sont commandés et répétables à l'identique. La section suivante présente la méthode d'acquisition des différents environnements et leur restitution permettant leur capture à travers une caméra événementielle virtuelle.

Le jeu de données 3D

Pouvant être généré sous un logiciel dédié, le jeu de données 3D permet l'exploration de scènes réelles capturées à l'aide d'une caméra virtuelle s'y déplaçant. Cette section décrit comment ces scènes réelles ont été capturées et comment une caméra peut s'y déplacer et générer des spikes compréhensibles pour le réseau de neurones.

Capture des scènes en 3D

Afin d'effectuer de la capture de scènes réelles et de les numériser pour leur exploitation à l'aide de logiciels, plusieurs technologies sont possibles aujourd'hui dont notamment celle utilisée ici, la technologie LiDAR (*Light Detection And Ranging*) ou lasergrammétrie. Cette dernière est basée sur la mesure de rebond de tirs de faisceaux lasers émis à intervalles réguliers. En mesurant le temps de trajet des faisceaux entre l'émetteur et le premier obstacle rencontré, il est possible d'en déterminer sa distance et d'ainsi obtenir un nuage de points tridimensionnel correspondant aux alentours de l'émetteur LiDAR. Ce nuage de points alors obtenu est une fidèle représentation spatiale et virtuelle d'une scène réelle.

Néanmoins, cette représentation ne présente pas de couleurs ou de formes précises pouvant être correctement interprétées. Pour pouvoir y appliquer

les formes, couleurs et textures, le nuage de point généré doit être converti en un maillage. Un maillage permet de reproduire un objet ou une surface par numérisation 3D par un arrangement de polygones, majoritairement des triangles. En appliquant un maillage au nuage de points généré par lasergrammétrie, il est obtenu une surface sur laquelle il est possible d'appliquer des textures et des couleurs. Toutefois, si l'on souhaite se rapprocher au mieux d'une scène réelle capturée il est préférable d'appliquer certains principes hérités de la photogrammétrie.

La photogrammétrie consiste à prendre des photographies d'une scène que l'on souhaite numériser et de l'appliquer à un maillage correspondant aux propriétés spatiales de l'environnement. Grâce à cette technique, il est également possible d'appliquer des photographies au maillage obtenu par la génération du nuage de points d'une acquisition LiDAR. Cette chaîne de traitement est représentée par la figure 5.1.

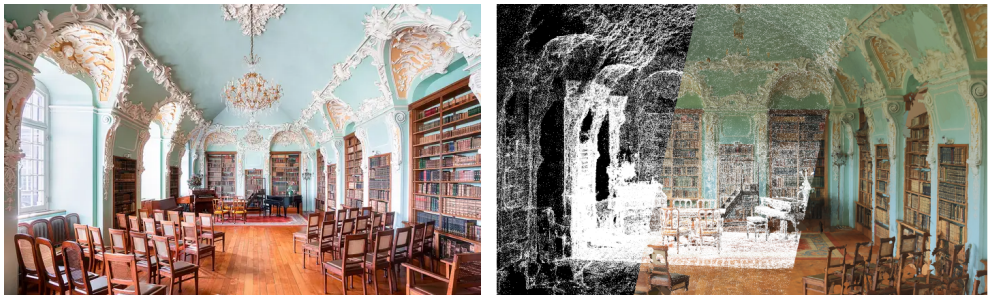


FIGURE 5.1 – Chaîne de traitement d'un environnement capturé par lasergrammétrie à sa numérisation tridimensionnelle par [Robroek, 2020]. A gauche : la scène réelle capturée. A droite : Les différentes étapes de traitement, de gauche à droite : Le nuage de points obtenu par acquisition LiDAR, la génération du maillage depuis le nuage de points et la colorisation des polygones par principe de photogrammétrie, la texturisation du maillage pour effacer les arêtes des polygones.

Pour la génération de ce jeu de données, différents environnements ont été capturés grâce aux capteurs LiDAR et photographique de l'iPad Pro 2020 d'Apple et l'application *Polycam - LiDAR & 3D Scanner*. L'acquisition des données LiDAR et photographiques s'effectuent en parallèle, en parcourant l'environnement à numériser (voir figure 5.2). Les informations

FIGURE 5.2 – Prévisualisation du maillage lors de l'acquisition des données LiDAR par l'application Polycam à l'aide du LiDAR intégré dans les appareils Apple.

inertielles de l'appareil sont aussi enregistrées pour chaque donnée afin de pouvoir faire correspondre les coordonnées des points aux photographies pour le maillage. Les nuages de points et les maillages obtenus après acquisition sont ensuite transférés vers un logiciel de rendu 3D dédié à leur exploration par une caméra virtuelle. Cette exploration permet alors la génération de spikes à la manière d'une caméra événementielle dont le fonctionnement est décrit ci-après. L'exploration d'une scène numérisée est illustrée par la figure 5.3.

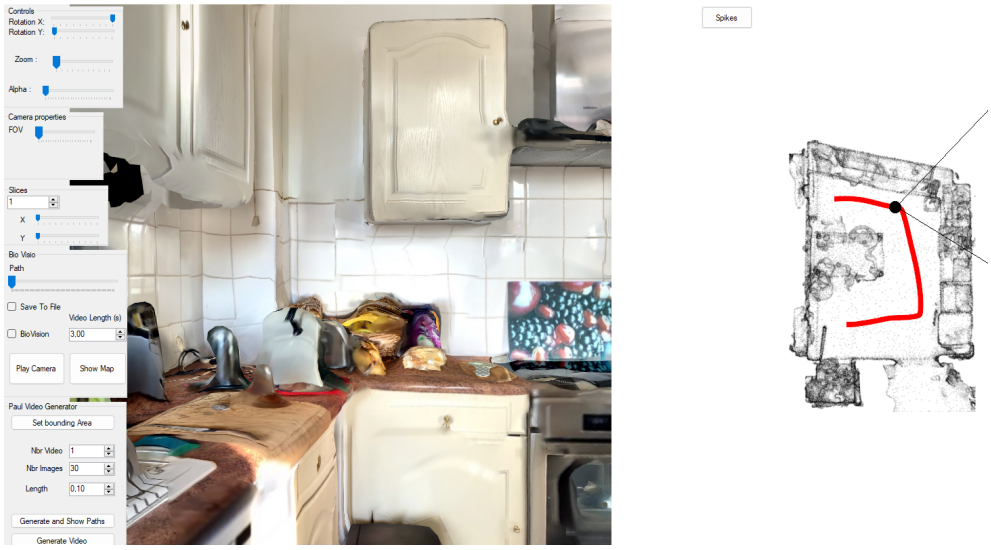


FIGURE 5.3 – Capture d'écran du logiciel dédié à l'exploration des environnements capturés. Sur la gauche est visible le rendu tridimensionnel de la scène capturée à partir du nuage de points visible sur la droite. Il est possible sur ce nuage de points de tracer un chemin que la caméra virtuelle doit parcourir. Il est également possible de définir une zone dans laquelle la caméra peut se déplacer selon différentes composantes. Le nombre d'images générées lors d'un suivi de tracé, le degré de champ de vision, la résolution des images capturées ainsi que le nombre d'images capturées par seconde sont tous des paramètres modifiables par l'utilisateur du logiciel.

Simulation de la caméra événementielle et génération des spikes

Lors d'une exploration, la caméra virtuelle suit un tracé soit défini à la main, soit généré automatiquement après une délimitation de la zone à explorer (tracé rouge sur la figure 5.3). En suivant le tracé, la caméra peut se déplacer de différentes façons afin de faire apparaître les différentes composantes du flux optique en ayant son champ de vision parallèle au tracé pour les composantes radiales, perpendiculaire pour les composantes translationnelles, ou en tournant sur elle-même pour des composantes rotationnelles. La figure 5.4 illustre ces différentes conditions.

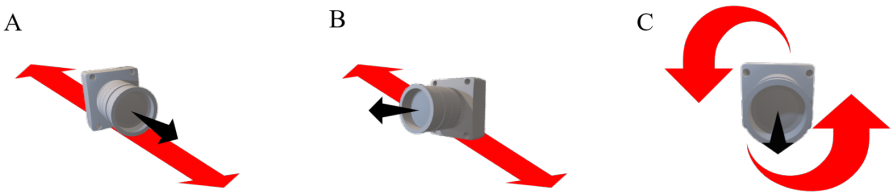


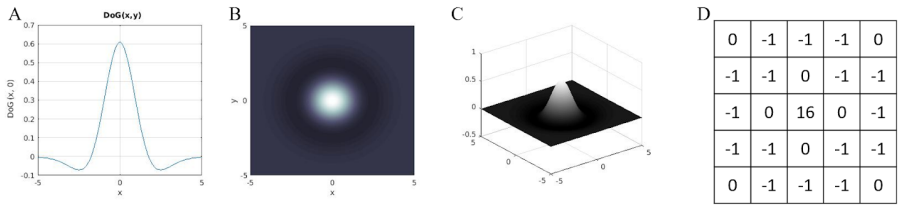
FIGURE 5.4 – Les différentes conditions de déplacement de la caméra pour la génération de spikes selon les composantes du flux optique (en rouge le chemin que suit la caméra, en noir la direction du champ de la caméra). A) Le champ de la caméra est orienté parallèlement au tracé et la caméra se déplace vers l'avant ou vers l'arrière pour la génération de composantes radiales. B) Le champ de la caméra est orienté perpendiculairement au tracé et la caméra se déplace suivant le tracé pour la génération de composantes translationnelles. C) La caméra tourne sur elle-même pour la génération de composantes rotationnelles.

Pendant tout le suivi du parcours, la caméra enregistre des séries d'images reflétant la scène capturée. Ces images sont ensuite traitées selon le principe de filtrage spatio-temporel utilisé par la caméra Yumain décrite en annexe A. Ce filtrage, bio-inspiré par le traitement rétinien chez le primate, est décrit ci-dessous.

Toutes les images capturées par la caméra virtuelle durant un suivi de tracé suivent plusieurs étapes de traitement pour en extraire des événements. La caméra capture les images à un taux de rafraîchissement pa-

ramétrable par l'utilisateur du logiciel. Aussi le temps mis par la caméra pour suivre le parcours tracé du point de départ au point d'arrivée rend la vitesse de la caméra variable selon ces deux paramètres, ainsi que la distance séparant les deux points. Les images ensuite obtenues sont traitées par paires : l'image capturée à un temps t et celle capturée à l'instant temporel suivant $t + 1$. Un filtre de convolution spatiale est appliqué à chaque image en utilisant un noyau DoG, ce type de filtrage étant observé dans la rétine comme décrit dans le chapitre 2 [Shapley and Enroth-Cugell, 1984]. Ce noyau, décrit par l'équation 5.1 permettant d'obtenir son profil spatial visible en figure 5.5, est le résultat d'une différence entre deux Gaussiennes de variances différentes.

$$DoG(x,y) = e^{\frac{-(x^2+y^2)}{2\sigma_1^2}} - \left(\frac{\sigma_1}{\sigma_2}\right)^2 \cdot e^{\frac{-(x^2+y^2)}{2\sigma_2^2}} \tag{5.1}$$



0	-1	-1	-1	0
-1	-1	0	-1	-1
-1	0	16	0	-1
-1	-1	0	-1	-1
0	-1	-1	-1	0

FIGURE 5.5 – Les différents profils spatiaux du noyau DoG pour le filtrage spatial des images de la caméra pour la génération des événements. A) Le profil spatial 1D de la fonction DoG avec $\frac{\sigma_2}{\sigma_1} = 1.6$. B-C) Les profils spatiaux 2D et 3D de cette même fonction DoG. D) Le kernel DoG équilibré utilisé pour le filtrage par convolution des images capturées par la caméra.

Une fois que ce filtrage spatial est effectué, la paire d'images successives subissent un filtrage temporelle par différenciation : l'image la plus récente est soustraite à la moins récente en fonction des intensités des pixels de l'image, ainsi une trace des changements entre les deux images à travers les différences de luminance. Cette opération laisse finalement apparaître les événements de type 'ON' et 'OFF' selon le signe de l'intensité lumineuse de chaque pixel. Les événements sont finalement générés selon la valeur de l'intensité lumineuse de chaque pixel selon le principe d'*intensity-to-latency*

conversion [Masquelier and Thorpe, 2010] où plus cette valeur est élevée, plus un événement arrive tôt. Cette génération permet finalement un traitement asynchrone des données capturées en répétant le même processus pour chaque paires d'images jusqu'à épuisement des captures. Les événements ainsi obtenus sont interprétables pour notre SNN et son apprentissage.

Architecture du réseau de neurones

L'architecture du SNN utilisé pour cette étude est essentiellement la même que celle utilisée au chapitre précédent. Le réseau est divisé en deux couches présentant chacune 64 neurones de type LIF. La connexion des données en entrées se fait de manière rétinotopique vers la première couche et de manière *fully-connected* entre la première et la seconde couche. Les données sont transmises sous forme de train de spikes sous le format AER avant de passer par un ordonnanceur traitant tous les événements du SNN. L'apprentissage du SNN se fait par la règle d'apprentissage non supervisée STDP pour la mise à jour des poids synaptiques des connexions, initialement aléatoires. Chaque couche présente également un mécanisme d'inhibition latérale. Les premiers résultats obtenus par l'utilisation du jeu de données 3D et du SNN sont présentés dans la section suivante.

5.2 Résultats

Ici sont présentés les premiers résultats obtenus grâce à l'apprentissage des composantes induites par les mouvements de caméra au sein d'environnements 3D.

Les composantes translationnelles

Le SNN est d'abord entraîné à l'aide de mouvements de translation générés au sein d'un environnement numérisé. Pour cela, la caméra virtuelle se déplace selon la condition présentée par la figure 5.4-B, son champ de vision est dirigé à la perpendiculaire du chemin tracé à suivre. Concernant

ces tracés, ils sont générés automatiquement au sein de la scène numérisée : une zone est délimitée sur le logiciel sur la vue en nuage de points dans laquelle la caméra est autorisée à effectuer des déplacements rectilignes en définissant aléatoirement un point de départ et un point d'arrivée. Cette génération automatique et aléatoire de tracés rectilignes pour le traitement des composantes translationnelles du flux optique permet une grande variation des propriétés spatiales des scènes capturées afin de pouvoir éviter un apprentissage qui engendrerait une sélectivité aux objets présents dans la scène plutôt qu'au mouvement perçu.

Le nombre de trajectoires générées pouvant être parcourues dans un sens ou dans l'autre permettant une extraction d'événements correspondant à des composantes translationnelles orientées vers la gauche ou vers la droite est fixé à 100, nombre de trajectoires similaires au premier jeu de données étudié. Un aperçu de deux trajectoires est illustré par la figure 5.6. Les événements extraits de ces vidéos de trajectoire grâce à la simulation de caméra événementielle sont ensuite envoyés en entrée du réseau pour leur apprentissage. Ainsi le SNN voit 90 trajectoires différentes lors des phases d'apprentissage et de classification et 10 sont réservées pour la phase de test. Les 90 trajectoires présentant les deux conditions de composantes translationnelles sont présentées uniformément dans un ordre aléatoire et sont toutes présentées trois fois afin de permettre au SNN de converger vers des champs récepteurs interprétables et capables de discriminer les composantes allant vers la droite et vers la gauche.

FIGURE 5.6 – Les images générées aux instants t_0, t_5, t_{10}, t_{15}, t_{20}, t_{25} et t_{30}, pendant le parcours de la caméra au sein de l'environnement numérisé selon les composantes translationnelles (A) vers la gauche et (B) vers la droite pour deux parcours différents.

Les champs récepteurs des neurones de la seconde couche du SNN obtenus après apprentissage des composantes translationnelles sont montrés par la figure 5.7. Ces derniers montrent alors une sélectivité aux composantes du flux optique selon les directions allant vers la droite et vers la gauche. Cette sélectivité s'affranchit des propriétés spatiales et statiques de la scène capturée (e.g., les objets, les textures, les surfaces) pour ne s'intéresser qu'aux mouvements de la caméra perçus.

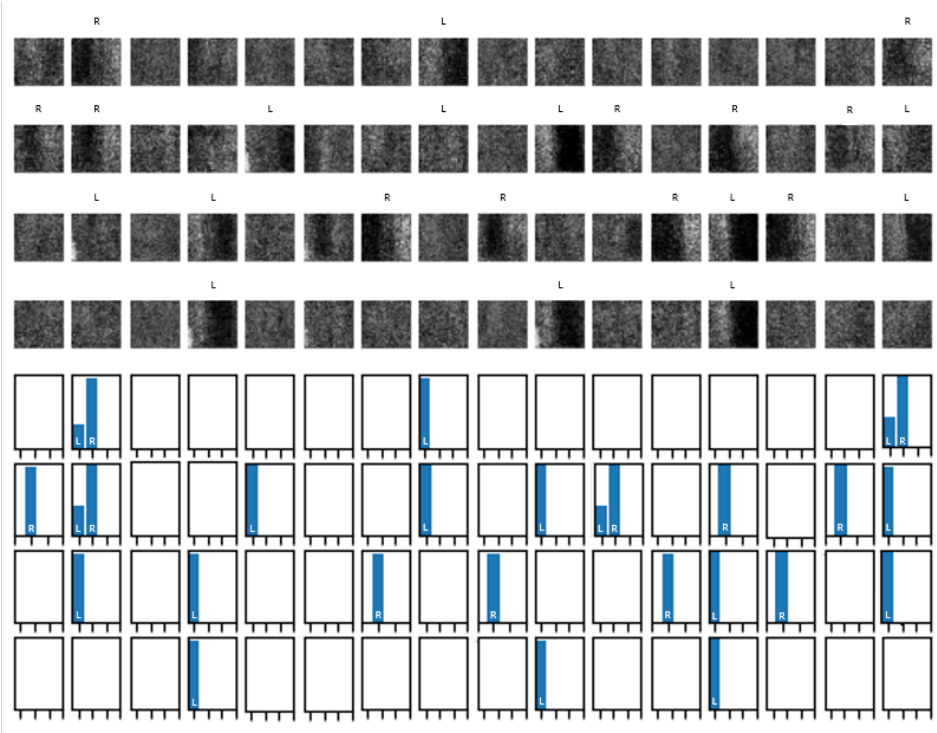


FIGURE 5.7 – Champs récepteurs obtenus après apprentissage sur des mouvements de translation générés à l'aide de la caméra virtuelle au sein d'un environnement numérisé. Les lettres 'L' et 'R' au-dessus des champs récepteurs correspondent au type de mouvement auquel ces derniers répondent. Ces lettres sont à mettre en corrélation avec la partie basse de la figure qui met en avant l'activité d'un neurone en fonction des mouvements présentés après apprentissage. Ici la lettre 'L' (pour *left*) correspond à la sélectivité aux composantes translationnelles allant vers la gauche, tandis que la lettre 'R' (pour *right*) correspond à la sélectivité aux composantes translationnelles allant vers la droite.

Cette première application du jeu de données 3D avec le SNN développé pour l'apprentissage des composantes du flux optique lors de la locomotion montre que cette dernière est possible pour la détection des composantes translationnelles. Afin de se rapprocher du jeu de données de navigation réelle utilisé dans le chapitre précédent, il est maintenant question d'ajouter à ces composantes translationnelles les composantes radiales du flux optique.

Les composantes radiales

La génération des composantes radiales au sein du logiciel intégrant les environnements numérisés s'effectue de la même manière que précédemment. Une zone est délimitée dans laquelle la caméra peut se déplacer selon des tracés rectilignes générés à partir d'un point de départ et d'un point d'arrivée aléatoires. Le champ de vision est cette fois parallèle au tracé, comme illustré par la figure 5.4-A, permettant alors l'apparition de composantes radiales du flux optique. Ces composantes sont alors décrites comme une expansion de la scène visuelle si la caméra avance, ou en contraction si la caméra recule. Les images générées par ces parcours sont représentées par la figure 5.8.

FIGURE 5.8 – Les images générées aux instants t_0, t_5, t_{10}, t_{15}, t_{20}, t_{25} et t_{30}, pendant le parcours de la caméra au sein de l'environnement numérisé selon les composantes radiales (A) en expansion et (B) en contraction pour deux parcours différents.

L'apprentissage s'effectue de la même manière que précédemment en présentant de manière répétée les différentes vidéos événementielles incluant des composantes radiales en plus des composantes translationnelles à partir

d'un état du réseau vierge de tout précédent apprentissage. La figure 5.9 montre l'organisation spatiale des champs récepteurs obtenus après convergence du réseau.

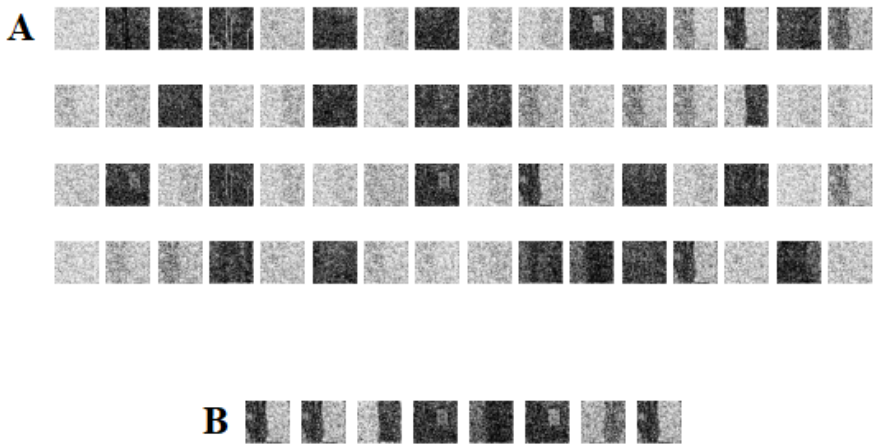


FIGURE 5.9 – A) Champs récepteurs obtenus après la phase d'apprentissage du réseau sur les composantes radiales et translationnelles. Peu de champs récepteurs ont finalement convergé vers une forme permettant l'interprétation concrète de leur sélectivité. Cependant certains peuvent attirer l'attention quant à leur organisation spatiale. B) Champs récepteurs remarquables. Ici est reconnaissable l'organisation spatiale des champs récepteurs des composantes radiales vue dans la figure 5.7. Une autre forme de champ récepteur se remarque également, montrant une zone centrale 'ON' et un pourtour 'OFF'. Ceci traduit alors une sélectivité aux composantes radiales et plus particulièrement à la condition de contraction du champ visuel.

Les champs récepteurs convergent alors plus difficilement vers une organisation spatiale convaincante et capable de discriminer correctement les deux types de composantes présentées. Cependant, il reste observable qu'une partie des neurones va en ce sens et confirme la possibilité du SNN présenté lors de cette étude et l'étude précédente à modéliser le traitement du flux optique et l'apprentissage de ses différentes composantes au cours de tâches de locomotion.

5.3 Conclusion et discussion

Ce chapitre a présenté la création d'un jeu de données consistant en des environnements numérisés pour leur exploration contrôlée à l'aide de caméra événementielle simulée. Les données événementielles générées ont été présentées au réseau de neurones impulsionnels décrit dans le chapitre précédent. Les premiers résultats obtenus suggèrent que les neurones du réseau artificiel deviennent progressivement sélectifs aux différentes composantes du flux optique dans des conditions de navigation simulées à partir d'environnements numérisés. Cette sélectivité est observable au niveau de la structure des champs récepteurs de neurones du réseau obtenus après apprentissage favorisant les composantes du flux optique et restant invariant aux propriétés spatiales des scènes capturées.

Concernant les environnements numérisés, leur exploration à l'aide de caméras virtuelles selon des chemins bien tracés constitue une première étape dans l'exploitation de ce jeu de données. Si les données événementielles générées à l'aide de cet environnement bien contrôlé constituent une excellente base pour l'apprentissage des composantes du flux optique, capturer des données dans des conditions de locomotion réelle pour tester les propriétés apprises par le réseau au sein des mêmes environnements se rapprocherait un peu plus d'une application de navigation réelle. Une des suites envisagées pour cette étude consiste à capturer les mouvements de la tête d'une personne se déplaçant au sein des environnements réels qui ont déjà été numérisés. Ceci permettrait de récupérer les mouvements naturels de la tête lors de tâches de navigation. Ces mouvements peuvent être capturés à l'aide de systèmes comme la motion capture (utilisée dans [Debat et al., 2021] par exemple) ou à l'aide des lunettes de réalité mixte HoloLens de Microsoft qui une fois sur la tête d'un participant est capable d'enregistrer tous les mouvements effectués (voir figure 5.10). Ces mouvements naturels pourraient alors être restitués au sein du logiciel dédié et la caméra pourrait alors suivre le chemin généré par les retranscriptions de ces mouvements, générant ainsi des données événementielles semblables à celles qui pourraient être retrouvées au cours de captations réelles. Une prochaine étape

sera de passer finalement à des captations réelles à l'aide d'une caméra événementielle montée sur la tête d'un participant lors de la locomotion.

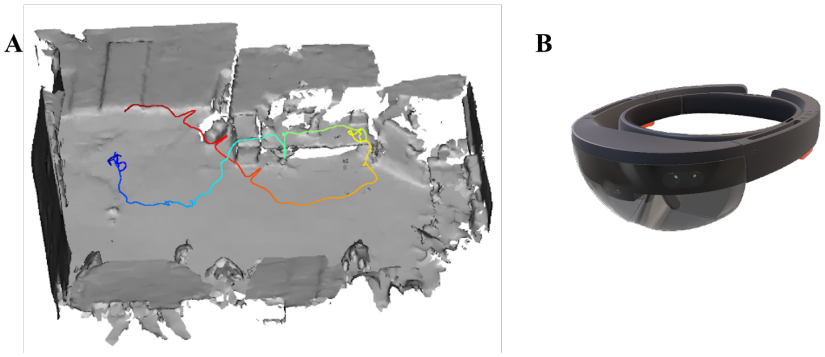


FIGURE 5.10 – A) Restitution des mouvements de la tête capturés par le système HoloLens de Microsoft lors de l'exploration d'une pièce. La pièce a été numérisée pour y représenter le trajet réellement parcouru et tracé en couleur, adapté de [Hübner et al., 2020]. B) Le système HoloLens de Microsoft présentant une multitude de capteurs permettant l'acquisition de données ainsi que la restitution des mouvements effectués lorsqu'il est porté.

Afin d'aller plus loin, il serait nécessaire de s'intéresser également aux composantes rotationnelles d'une part, ainsi que leurs combinaisons avec les composantes radiales d'autre part. Ces combinaisons constituent des composantes de flux optique en spirale qui traduisent le mouvement de soi sur un trajet curvilinéaire, et dépendent en conditions naturelles de l'orientation du regard pour l'appréciation du niveau de courbure d'un chemin à parcourir. Les récentes études de [Layton and Fajen, 2022] considèrent ces composantes à travers un modèle de l'aire MSTd et concluent sur la présence de neurones spécifiquement sélectifs à cette nature de composantes, montrant des sorties plus précises que des neurones sélectifs aux composantes radiales ou rotationnelles seulement.

Les travaux présentés dans ce chapitre et incluant la numérisation d'environnements, leur exploration par une caméra événementielle virtuelle pour la génération d'événements, le tout au sein d'un logiciel 3D dédié, ainsi que l'apprentissage de ces données par un réseau de neurones impulsionnels font

actuellement l'objet de la rédaction d'un article qui sera prochainement soumis pour publication.

Chapitre 6

Discussion générale

Tout au long de cette thèse et à travers les différents chapitres de ce manuscrit, je me suis intéressé à la modélisation du traitement visuel du flux optique lors de la locomotion, essentiellement lors de tâches de *heading*. Je me suis notamment basé sur le traitement de ce flux optique chez le primate qui a été abondamment documenté à partir d'exploration en électrophysiologie effectuées au sein des aires MT et MST du primate [Sato et al., 2010, Takahashi et al., 2007, Gu et al., 2010, Warren and Rushton, 2009, Fajen et al., 2013], aires mises en avant par les figures 2.3 et 2.7. J'ai pu proposer un modèle computationnel bio-inspiré qui se base sur des réseaux des neurones impulsionnels qui reçoivent en entrée des données événementielles simulées ou bien capturées par des caméras asynchrones. Ces réseaux sont composés de neurones intégrateurs à fuite (neurones LIF) et régulés par une loi d'apprentissage non supervisée de type STDP décrits par la section 3.2. Ce chapitre revient sur les résultats obtenus avec ce modèle à travers l'apprentissage de différents jeux de données événementiels, et les perspectives qu'amènent ces résultats pour de futures études.

6.1 Résumé des résultats

Une première étude, développée dans le chapitre 4, a montré que les neurones d'un réseau impulsionnel entraînés à l'aide d'une règle d'apprentis-

sage STDP non-supervisée, deviennent progressivement sélectifs aux composantes du flux optique lorsque le réseau reçoit en entrée des données simulées ou capturées par une caméra événementielle dans un contexte de navigation. Ainsi l'apprentissage sur des simulations constituées de mouvements naturalistiques, proposés par les déplacements des disques blancs sur fond noir (voir figure 4.1), a pu mettre en évidence l'évolution des propriétés spatiales des champs récepteurs des neurones. Cela a également pu montrer leur capacité à discriminer les différentes composantes de flux optique, qu'elles soient translationnelles, rotationnelles ou radiales, validant ainsi l'architecture du SNN proposé avec des connexions rétinotopiques au sein de la première couche et *fully connected* par la suite. Cette architecture s'est ensuite vue être appliquée à un jeu de données plus complexe présentant une tâche de locomotion naturelle au sein d'un environnement urbain capturé à l'aide d'une caméra événementielle tenue par un piéton [Mueggler et al., 2017]. Si le réseau arrive à converger, après entraînement, vers une sélectivité aux différentes composantes de flux optique induite par le mouvement du piéton tenant la caméra événementielle, celle-ci n'est pas parfaite, montrant une confusion entre les composantes translationnelles allant vers la droite et les autres composantes présentes, malgré une précision de 87.5 pourcent, ce qui peut-être expliqué par différents facteurs. Tout d'abord le niveau de complexité est beaucoup plus élevé par rapport au jeu de données précédent qu'il s'agisse d'un environnement ou de mouvements beaucoup plus complexes. Les données événementielles générées par la caméra sont bruitées et sont très fortement affectées par les mouvements brusques et parasites de la tête du piéton offrant un contexte d'apprentissage pour le SNN relativement délicat. Ensuite, le chemin parcouru avec la caméra présente des conditions de navigation et d'orientation très hétérogènes. Le jeu de données est prédominé par les composantes translationnelles allant vers la gauche et par les composantes radiales d'expansion de la scène visuelle, délaissant par exemple les composantes translationnelles allant vers la droite qui ne sont représentées que très faiblement par rapport aux deux autres. Finalement, le manque de données contrôlées et de données *ground-truth* rendent ce jeu de données trop complexe pour des analyses plus poussées,

malgré les capacités du réseau à identifier et discriminer les composantes de flux optique dans ce cas. Ces constatations ont mené à la création d'un jeu de données à mi-chemin entre les données événementielles simulées et les données réelles captées à l'aide d'une caméra événementielle.

La seconde étude, sujet du chapitre 5, se penche sur la création d'un tel jeu de données. Pour ce faire, différents environnements sont numérisés à l'aide de la lasergrammétrie. Cette méthode permet de restituer les propriétés spatiales d'un environnement dans un format 3D numérique. A l'aide d'un logiciel dédié, les environnements numérisés peuvent être explorés à partir du déplacement d'une caméra virtuelle simulant le comportement d'une caméra événementielle. Cette exploration parfaitement contrôlée permet de générer des événements en fonction des mouvements de la caméra permettant l'extraction des différentes composantes du flux optique. Ce jeu de données décrit finalement un environnement réel dans lequel se déplace virtuellement une caméra simulant un fonctionnement événementiel, à mi-chemin entre la simulation et la prise de vue réelle. Le même SNN est ensuite utilisé pour traiter les données générées par l'exploration des environnements. Je me suis tout d'abord intéressé à un apprentissage basé sur des mouvements translationnels. Les neurones une fois entraînés avec ces données deviennent sélectifs aux deux types de mouvement étudiés, bien que les scènes présentées en entrée étaient extrêmement variées. Ces premiers résultats permettent de valider le modèle quant à sa capacité à traiter le flux optique tout en étant invariant aux propriétés spatiales des scènes statiques comme les objets ou les textures, marqueur de la faculté d'adaptabilité des règles d'apprentissage non supervisées à tout type d'environnements. Dans un second temps sont ajoutées les composantes radiales du flux optique en plus des translationnelles. La convergence du réseau s'avère dans ce cas là plus difficile mais fait néanmoins apparaître des structures remarquables au sein des champs récepteurs des neurones du SNN. Ce jeu de données 3D constitue alors une première base convaincante pour l'apprentissage des informations visuelles lors de la locomotion et l'entraînement de SNNs pour ces tâches.

6.2 Perspectives

Les réseaux de neurones impulsionnels bio-inspirés que j'ai développés au cours de mon doctorat sont donc capables de modéliser le traitement visuel des informations lors de la locomotion et l'émergence de la sélectivité aux composantes de flux optique. Néanmoins, le traitement du flux optique et son étude a pu mettre en lumière deux mécanismes décrits dans le chapitre 2 : le *heading* et le *flow parsing*, ceux-ci désignent respectivement la direction vers laquelle un observateur se dirige, et la capacité d'un observateur à estimer sa trajectoire et ses déplacements en fonction d'éléments extérieurs en mouvements. Lors des travaux présentés, c'est le mécanisme de *heading* qui a été mis en avant, traduisant des conditions de navigation orientées vers une direction précise sans interférence de l'environnement. Il est alors envisageable de considérer le *flow parsing* en plus des tâches de *heading* étudiées à l'aide de SNNs.

Heading

La modélisation du *heading* a été l'objet des différentes études présentées ici et de l'apprentissage du réseau de neurones impulsionnels. Ce réseau a pu montrer sa capacité à apprendre les différentes composantes du flux optique dans un contexte de navigation contextuelle orientée vers une direction précise que celle-ci s'effectue en translation, rotation, ou en expansion / contraction de la scène visuelle. Ainsi l'apprentissage effectué a pu être catégorisé et classifié grâce à la méthode décrite par [Diehl and Cook, 2015] reposant sur l'activité des neurones en fonction des composantes présentées. D'autres méthodes pourraient être envisagées pour cette tâche. Par exemple, [Debat et al., 2021] utilisent une régression polynomiale afin de déterminer la trajectoire d'une balle en mouvement. D'autres méthodes de classification seraient également envisageables à partir des configurations spatiales présentées par les champs récepteurs des neurones après entraînement. Ainsi, ces derniers pouvant être traités comme des images, il serait possible de leur appliquer des algorithmes de *clustering* comme le Principal

Component Analysis (PCA) [Pearson, 1901] ou l'algorithme t-SNE [Maaten and Hinton, 2008] par exemple.

Flow parsing

Cette modélisation du *flow parsing* a déjà pu faire l'objet d'étude comme celle de [Layton et al., 2012] présentée au chapitre 2, en reprenant des données biologiques collectées par [Royden and Hildreth, 1996] et [Warren and Saunders, 1995] afin d'établir un modèle computationnel. Cependant, cette modélisation pour l'aide à la locomotion grâce au *flow parsing* n'inclut pas l'utilisation d'un SNN. Cela nécessiterait alors de pouvoir évaluer la profondeur des scènes observées, ainsi que de prédire le parcours d'un objet en mouvement, puisque le mécanisme de *flow parsing* consiste à appréhender le mouvement des objets se déplaçant dans notre champ visuel, que ceux-ci se déplacent vers nous ou non, et d'ajuster alors notre trajectoire afin de les éviter ou de les intercepter. Concernant l'évaluation de la profondeur à l'aide de SNNs, plusieurs études s'y sont penchés, notamment au sein de l'équipe SV3M du laboratoire CerCo avec une étude utilisant un réseau de neurones impulsionnels pour traiter les spikes générés à partir d'une base de données stéréoscopiques collectées au sein de scènes naturelles [Chauhan et al., 2018]. Cette étude a montré qu'un tel réseau est capable de devenir sélectif aux disparités binoculaires horizontales. A partir des réponses des neurones de ce réseau, il est donc possible d'estimer la profondeur des objets de la scène.

Une autre étude conduite au sein du laboratoire est celle de [Debat et al., 2021]. Cette étude s'intéresse à la prédiction de trajectoires à l'aide de caméras événementielles et d'un SNN doté d'une règle d'apprentissage STDP. Cette dernière montre l'efficacité des SNNs pour des tâches de suivi et de prédiction de mouvement, surpassant des prédictions humaines sur les mêmes tâches. Elle montre aussi des neurones témoignant d'une sélectivité à la direction d'une part, mais aussi à la vitesse d'autre part, indispensable pour une estimation de trajectoire correcte d'un objet. Cette sélectivité à la vitesse est obtenue par l'utilisation de différents délais synaptiques entre

les couches de neurones du SNN.

Délais

Pour permettre une sélectivité des neurones à la vitesse, il faut que ceux-ci puissent devenir sélectifs à la temporalité des événements leur parvenant. Ainsi, l'ajout de délais de transmission synaptique entre les neurones permet de générer un retard entre l'instant d'émission d'un événement et sa réception. Si l'on multiplie ces délais, on multiplie les événements générés pour un même mouvement, les neurones apprennent ces différentes propriétés temporelles pour un même mouvement, et permettent finalement de différencier des mouvements de différentes vitesses selon la fréquence d'activation des neurones concernés. Plusieurs méthodes existent pour l'ajout de délais synaptiques au sein de SNN. [Orchard et al., 2013] décrivent une approche ramenant tous les événements générés à être transmis en même temps, ralentissant alors leur transmission en fonction du temps mis par le dernier événement a être généré sur la fenêtre temporel étudiée (voir figure 6.1-A). Une seconde méthode est celle proposée par [Paredes-Vallés et al., 2020]. Cette dernière consiste à multiplier les connexions synaptiques entre deux neurones, chacune ayant un délai défini retardant un même événement autant de fois qu'il y a de synapses (voir figure 6.1-B).

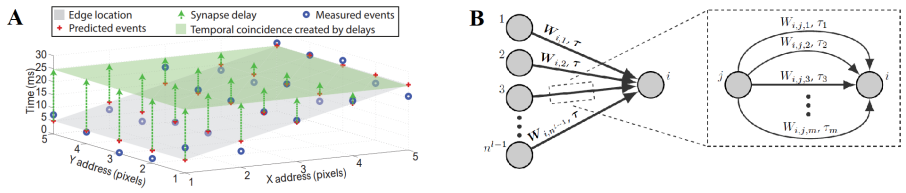


FIGURE 6.1 – A) La méthode de génération de délais employée par [Orchard et al., 2013] consistant à faire coïncider l'arrivée des événements d'une même fenêtre temporelle au même instant. B) La méthode de génération de délais synaptiques employée par [Paredes-Vallés et al., 2020] en multipliant les synapses entre deux neurones pour autant de délais différents souhaités.

L'ajout de délais synaptiques au SNN présenté dans les précédents chapitres de ce manuscrit pourrait alors permettre aux neurones de devenir

sélectifs à différentes vitesses. Cela pourrait ainsi constituer un premier pas vers l'intégration du mécanisme de *flow parsing* puisque cela permettrait d'évaluer la vitesse à laquelle un observateur se déplace ainsi que la vitesse des objets qui l'entourent. Cette vitesse perçue ne dépend néanmoins pas seulement de la vitesse de déplacement puisque cette perception est également différente selon l'excentricité rétinienne, c'est-à-dire si les informations sont plutôt projetées sur la partie centrale du champ de vision ou sur sa périphérie.

Vision centrale et périphérique

Cette division entre vision centrale et périphérique émane notamment de la répartition des photorécepteurs sur la rétine. Ainsi la vision centrale correspond à la zone avec la plus forte concentration de cônes et code 2 degrés du champ de vision total. Elle permet de voir avec précision. La vision périphérique, correspondant aux zones où les bâtonnets sont en majorité et décrit tout ce qui n'est pas en vision centrale. C'est cette vision qui nous informe sur l'environnement dans lequel nous nous trouvons et permet une meilleure appréciation des mouvements que la vision centrale mais en étant moins précise. L'intérêt de pouvoir reproduire ce mécanisme biologique et de le modéliser à travers l'organisation des connexions entre l'entrée d'un réseau de neurones et sa première couche serait de se rapprocher encore un peu plus du vivant. Au sein de SNN, cela permettrait l'obtention de filtres et champs récepteurs différents selon leur position attribuée par les connexions adéquates. Il est alors possible d' imaginer que des neurones connectés au centre du champ visuel convergeront vers une sélectivité aux fréquences spatiales élevées et aux vitesses réduites, tandis que ceux connectés à la périphérie du champ visuel seraient à l'inverse plus enclin à devenir sélectifs aux fréquences spatiales faibles et aux vitesses élevées, la vitesse du flux étant directement proportionnelle à l'excentricité rétinienne dans des conditions naturelles de navigation.

Cette sélectivité au flux optique par la vision périphérique à d'ailleurs pu être mis en évidence chez l'humain grâce à une étude menée au sein du

laboratoire et à laquelle j'ai pu participer [Guénot et al., 2022]. Cette étude s'est intéressée à la détection du flux optique chez des patients atteints de dégénérescence maculaire liée à l'âge (DMLA) qui engendre l'apparition d'un scotome obstruant la vision centrale. Les participants de cette étude, des personnes atteintes de DMLA et des personnes contrôles appareillées en âge sans trouble de la vision, ont été confrontés à des tâches de discrimination du flux optique selon les composantes translationnelles, rotationnelles et radiales. Ce flux optique a été simulé à partir de nuages de points se déplaçant en fonction des composantes voulues en y incluant un paramètre de bruit s'adaptant aux réponses des participants. Il s'est avéré que les patients souffrant d'une dégénérescence maculaire sont meilleurs pour la détection de direction de mouvement du flux optique que les autres et que sa perception n'est pas détériorée. Les résultats obtenus mettent alors en avant que la détection du flux optique s'opère très largement par la vision périphérique comme ont pu le montrer d'autres études [Brandt et al., 1973, Bugnariu and Fung, 2007]. Il est également discuté de la performance des patients ayant perdu leur vision centrale et le temps entre l'étude et leur diagnostic. Ainsi il est noté une corrélation négative significative entre ce temps et les performances : plus le diagnostic est récent, moins les performances sont bonnes et inversement. Cela indiquerait que la faculté d'identification des différentes composantes du flux s'améliore avec le temps, amenant l'idée qu'une réorganisation corticale intervient après la perte de la vision centrale.

Modéliser cette réorganisation corticale serait également envisageable à l'aide de réseaux impulsionnels. Si un tel réseau est capable comme montré ici de modéliser le traitement du flux optique, et de s'adapter à différents environnements et conditions de présentation de données en entrées, cette grande adaptabilité pourrait également être mise à profit pour une meilleure compréhension des opérations et traitements effectués lors d'une réorganisation corticale témoignant de la plasticité du cerveau.

Améliorations et perspectives futures

Finalement, l'entraînement de réseaux de neurones impulsionnels est envisageable pour différentes modélisations du traitement visuel. Ces derniers peuvent servir à la détection de propriétés spatiales locales utile pour la détection de formes [Barbier et al., 2021] ou de visages [Mermillod et al., 2010, Entzmann et al., 2022]. Ces réseaux peuvent également s'intéresser à des tâches de détection au sein de scènes en mouvement plus globales comme pour la navigation autonome [Cordone et al., 2022] où pouvoir estimer le mouvement égocentré et détecter les obstacles relèvent un enjeu crucial de sécurité [Bak et al., 2010, Bak et al., 2014]. Les SNN peuvent également apporter au domaine de la vision prothétique et pour l'aide aux personnes aveugles, cadre d'étude du projet INCA dans lequel s'inscrit cette thèse. En effet, leur interfaçage avec des caméras événementielles et leur traitement bio-inspirés peuvent permettre une restitution de l'information de mouvement lors de la navigation traité à travers des implants rétiniens présentant une faible résolution [Desvergnes et al., 2022]. Ainsi toute restitution se doit d'être optimisée afin de rester compréhensible et des SNNs bio-inspirés pourraient alors relever ce défi.

De tels réseaux pourraient profiter de tous les mécanismes discutés ici, qui constituent ainsi d'intéressantes directions pour leur amélioration. Cependant, il est bon de rappeler que le réseau présenté tout au long de ce manuscrit repose sur des mécanismes biologiques inspirés du traitement visuel chez l'humain. Aller plus loin pourrait alors nous éloigner de cette bio-inspiration initiale qui peut souffrir de ses performances, mais qui constitue un excellent cadre pour la modélisation des différents traitements s'effectuant au sein des voies visuelles corticales.

Bibliographie

[Aamir et al., 2018] Aamir, S. A., Stradmann, Y., Müller, P., Pehle, C., Hartel, A., Grübl, A., Schemmel, J., and Meier, K. (2018). An Accelerated LIF Neuronal Network Array for a Large-Scale Mixed-Signal Neuromorphic Architecture. *IEEE Transactions on Circuits and Systems I : Regular Papers*, 65(12) :4299–4312.

[Abbott, 1999] Abbott, L. F. (1999). Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Research Bulletin*, 50(5-6) :303–304.

[Adelson and Bergen, 1985] Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America. A, Optics and Image Science*, 2(2) :284–299.

[Akopyan et al., 2015] Akopyan, F., Sawada, J., Cassidy, A., Alvarez-Icaza, R., Arthur, J., Merolla, P., Imam, N., Nakamura, Y., Datta, P., Nam, G.-J., Taba, B., Beakes, M., Brezzo, B., Kuang, J. B., Manohar, R., Risk, W. P., Jackson, B., and Modha, D. S. (2015). TrueNorth : Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34(10) :1537–1557.

[al Haytham et al., 1572] al Haytham, I., Witelo, and Risner, F. (1572). *Opticae thesaurus. Alhazeni Arabis libri septem, nunc primum editi :; eiusdem liber de crepusculis et nubium ascensionibus. item Vitelloni Thuringopoloni opticae libri X.* per Episcopios, Basileae. Accession Number : 990020167960205503 Source : ETH-BIB.

[Asghar et al., 2021] Asghar, M. S., Arslan, S., and Kim, H. (2021). A Low-Power Spiking Neural Network Chip Based on a Compact LIF Neuron and Binary Exponential Charge Injector Synapse Circuits. *Sensors*, 21(13) :4462. Publisher : MDPI AG.

[Bak et al., 2010] Bak, A., Bouchafa, S., and Aubert, D. (2010). Detection of independently moving objects through stereo vision and ego-motion extraction. In *2010 IEEE Intelligent Vehicles Symposium*, pages 863–870. ISSN : 1931-0587.

[Bak et al., 2014] Bak, A., Bouchafa, S., and Aubert, D. (2014). Dynamic objects detection through visual odometry and stereo-vision : a study of inaccuracy and improvement sources. *Machine Vision and Applications*, 25(3) :681–697.

[Barbier et al., 2021] Barbier, T., Teuliere, C., and Triesch, J. (2021). *Spike timing-based unsupervised learning of orientation, disparity, and motion representations in a spiking neural network.*

[Barlow, 1961] Barlow, H. (1961). Possible Principles Underlying the Transformations of Sensory Messages. *Sensory Communication*, 1.

[Beaubert et al., 2005] Beaubert, E., Pariguet, F., and Taboulot, S. (2005). *Manuel de l'opticien.* Maloine.

[Behnke, 2003] Behnke, S. (2003). Neurobiological Background. In Behnke, S., editor, *Hierarchical Neural Networks for Image Interpretation*, Lecture Notes in Computer Science, pages 17–33. Springer, Berlin, Heidelberg.

[Beyeler et al., 2016] Beyeler, M., Dutt, N., and Krichmar, J. L. (2016). 3D Visual Response Properties of MSTd Emerge from an Efficient, Sparse Population Code. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 36(32) :8399–8415.

[Bi and Poo, 1998] Bi, G.-q. and Poo, M.-m. (1998). Synaptic Modifications in Cultured Hippocampal Neurons : Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. *Journal of Neuroscience*, 18(24) :10464–10472. Publisher : Society for Neuroscience Section : ARTICLE.

[Bichler et al., 2012] Bichler, O., Querlioz, D., Thorpe, S. J., Bourgoin, J.-P., and Gamrat, C. (2012). Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity. *Neural Networks*, 32 :339–348.

[Bienenstock et al., 1982] Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity : orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 2(1) :32–48.

[Born and Bradley, 2005] Born, R. T. and Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, 28 :157–189.

[Boyd and Casagrande, 1999] Boyd, J. D. and Casagrande, V. A. (1999). Relationships between cytochrome oxidase (CO) blobs in primate primary visual cortex (V1) and the distribution of neurons projecting to the middle temporal area (MT). *Journal of Comparative Neurology*, 409(4) :573–591. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291096-9861%2819990712%29409%3A4%3C573%3A%3AAID-CNE5%3E3.0.CO%3B2-R.

[Brandli et al., 2014] Brandli, C., Berner, R., Yang, M., Liu, S.-C., and Delbruck, T. (2014). A 240 × 180 130 dB 3 μs Latency Global Shutter Spatiotemporal Vision Sensor. *IEEE Journal of Solid-State Circuits*, 49(10) :2333–2341.

[Brandt et al., 1973] Brandt, T., Dichgans, J., and Koenig, E. (1973). Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research*, 16(5) :476–491.

[Brenner et al., 2000] Brenner, N., Bialek, W., and Steveninck, R. d. R. v. (2000). Adaptive Rescaling Maximizes Information Transmission. *Neuron*, 26(3) :695–702. Publisher : Elsevier.

[Briggs and Usrey, 2011] Briggs, F. and Usrey, W. M. (2011). Corticogeniculate feedback and visual processing in the primate. *The Journal of Physiology*, 589(Pt 1) :33–40.

[Britten et al., 1992] Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion : a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, 12(12) :4745–4765. Publisher : Society for Neuroscience Section : Articles.

[Brooks et al., 2011] Brooks, K. R., Morris, T., and Thompson, P. (2011). Contrast and stimulus complexity moderate the relationship between spatial frequency and perceived speed : Implications for MT models of speed perception. *Journal of Vision*, 11(14) :19.

[Bruss and Horn, 1983] Bruss, A. R. and Horn, B. K. P. (1983). Passive navigation. *Computer Vision, Graphics, and Image Processing*, 21(1) :3–20.

[Bugnariu and Fung, 2007] Bugnariu, N. and Fung, J. (2007). Aging and selective sensorimotor strategies in the regulation of upright balance. *Journal of NeuroEngineering and Rehabilitation*, 4 :19.

[Burbank, 2015] Burbank, K. S. (2015). Mirrored STDP Implements Autoencoder Learning in a Network of Spiking Neurons. *PLOS Computational Biology*, 11(12) :e1004566. Publisher : Public Library of Science.

[Calvert, 1954] Calvert, E. S. (1954). Visual Judgments in Motion. *The Journal of Navigation*, 7(3) :233–251. Publisher : Cambridge University Press.

[Censi et al., 2015] Censi, A., Mueller, E., Frazzoli, E., and Soatto, S. (2015). A Power-Performance Approach to Comparing Sensor Families, with application to comparing neuromorphic to traditional vision sensors. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3319–3326. ISSN : 1050-4729.

[Chandradevan, 2017] Chandradevan, R. (2017). Radial Basis Functions Neural Networks — All we need to know.

[Chauhan et al., 2018] Chauhan, T., Masquelier, T., Montlibert, A., and Cottereau, B. R. (2018). Emergence of Binocular Disparity Selectivity through Hebbian Learning. *Journal of Neuroscience*, 38(44) :9563–9578.

[Chen et al., 2013] Chen, G., King, J. A., Burgess, N., and O'Keefe, J. (2013). How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1) :378–383.

[Clark et al., 2018] Clark, M. A., Douglas, M., and Choi, J. (2018). *Biolgy 2e*. OpenStax, Houston, Texas.

[Clemens et al., 2012] Clemens, J. M., Ritter, N. J., Roy, A., Miller, J. M., and Van Hooser, S. D. (2012). The laminar development of direction selectivity in ferret visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 32(50) :18177–18185.

[Cordone et al., 2022] Cordone, L., Miramond, B., and Thierion, P. (2022). Object Detection with Spiking Neural Networks on Automotive Event Data. arXiv :2205.04339 [cs].

[Crochet and Petersen, 2006] Crochet, S. and Petersen, C. C. H. (2006). Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nature Neuroscience*, 9(5) :608–610.

[Crook et al., 1988] Crook, J. M., Lange-Malecki, B., Lee, B. B., and Valberg, A. (1988). Visual resolution of macaque retinal ganglion cells. *The Journal of Physiology*, 396(1) :205–224. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1988.sp016959.

[Cutting and Readinger, 2002] Cutting, J. E. and Readinger, W. O. (2002). Perceiving motion while moving : How pairwise nominal invariants make optical flow cohere. *Journal of Experimental Psychology : Human Perception and Performance*, 28(3) :731–747.

[Cutting et al., 1992] Cutting, J. E., Springer, K., Braren, P. A., and Johnson, S. H. (1992). Wayfinding on foot from information in retinal, not optical, flow. *Journal of Experimental Psychology : General*, 121(1) :41–72. Place : US Publisher : American Psychological Association.

[Dacey, 2004] Dacey, D. (2004). 20 Origins of Perception : Retinal Ganglion Cell Diversity and the Creation of Parallel Visual Pathways. In Gazzaniga, M. S., editor, *The Cognitive Neurosciences Iii*, page 281. MIT Press.

[Dacey, 2000] Dacey, D. M. (2000). Parallel pathways for spectral coding in primate retina. *Annual Review of Neuroscience*, 23 :743–775.

[Dan and Poo, 2004] Dan, Y. and Poo, M.-M. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron*, 44(1) :23–30.

[Davies et al., 2018] Davies, M., Srinivasa, N., Lin, T.-H., Chinya, G., Cao, Y., Choday, S. H., Dimou, G., Joshi, P., Imam, N., Jain, S., Liao, Y., Lin, C.-K., Lines, A., Liu, R., Mathaikutty, D., McCoy, S., Paul, A., Tse, J., Venkataramanan, G., Weng, Y.-H., Wild, A., Yang, Y., and Wang, H. (2018). Loihi : A Neuromorphic Manycore Processor with On-Chip Learning. *IEEE Micro*, 38(1) :82–99.

[Dayan and Abbott, 2001] Dayan, P. and Abbott, L. F. (2001). *Theoretical neuroscience : computational and mathematical modeling of neural systems*. Computational neuroscience. Massachusetts Institute of Technology Press, Cambridge, Mass.

[Debat, 2021] Debat, G. (2021). *Extraction d'informations spatio-temporelles du mouvement avec un réseau de neurones à spike : application sur une tâche de prédiction de trajectoire de balles*. PhD thesis, Toulouse 3 - Paul Sabatier, Toulouse.

[Debat et al., 2021] Debat, G., Chauhan, T., Cottereau, B. R., Masquelier, T., Paindavoine, M., and Baures, R. (2021). Event-Based Trajectory Prediction Using Spiking Neural Networks. *Frontiers in Computational Neuroscience*, 15. Publisher : Frontiers.

[Delbruck, 2016] Delbruck, T. (2016). Neuromorophic vision sensing and processing. In *2016 46th European Solid-State Device Research Conference (ESSDERC)*, pages 7–14. ISSN : 2378-6558.

[Desvergnes et al., 2022] Desvergnes, J., Carlier, A., Ooi, W. T., Charvillat, V., and Jouffrais, C. (2022). Does Switching between Different Ren-

derings Allow Blind People with Visual Neuroprostheses to Better Perceive the Environment ? *ICCHP-AAATE 2022 Open Access Compendium "Assistive Technology, Accessibility and (e)Inclusion" Part I.* ISBN : 9783950499780.

[DeYoe and Van Essen, 1985] DeYoe, E. A. and Van Essen, D. C. (1985). Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature*, 317(6032) :58–61.

[Diamond, 2017] Diamond, J. S. (2017). Inhibitory Interneurons in the Retina : Types, Circuitry, and Function. *Annual Review of Vision Science*, 3 :1–24.

[Diehl and Cook, 2015] Diehl, P. U. and Cook, M. (2015). Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Frontiers in Computational Neuroscience*, 9. Publisher : Frontiers.

[Dokka et al., 2015] Dokka, K., MacNeilage, P. R., DeAngelis, G. C., and Angelaki, D. E. (2015). Multisensory Self-Motion Compensation During Object Trajectory Judgments. *Cerebral Cortex*, 25(3) :619–630.

[Duffy and Wurtz, 1991] Duffy, C. J. and Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65(6) :1329–1345.

[Duffy and Wurtz, 1995] Duffy, C. J. and Wurtz, R. H. (1995). Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 15(7 Pt 2) :5192–5208.

[Dyde and Harris, 2008] Dyde, R. T. and Harris, L. R. (2008). The influence of retinal and extra-retinal motion cues on perceived object motion during self-motion. *Journal of Vision*, 8(14) :5.

[Enroth-Cugell and Robson, 1966] Enroth-Cugell, C. and Robson, J. (1966). The contrast sensitivity of retinal ganglion cells on the cat. *J. Physiol. (Lond.)*, 187 :516–552.

[Entzmann et al., 2022] Entzmann, L., Guyader, N., Kauffmann, L., Peyrin, C., and Mermillod, M. (2022). Detection of emotional faces : the role of spatial frequencies and local features.

[Fajen et al., 2013] Fajen, B. R., Parade, M. S., and Matthis, J. S. (2013). Humans perceive object motion in world coordinates during obstacle avoidance. *Journal of Vision*, 13(8) :25.

[Falez, 2019] Falez, P. (2019). *Improving Spiking Neural Networks Trained with Spike Timing Dependent Plasticity for Image Recognition*. Theses, Université de Lille.

[Feldman, 2012] Feldman, D. E. (2012). The spike timing dependence of plasticity. *Neuron*, 75(4) :556–571.

[Felleman and Van Essen, 1991] Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, 1(1) :1–47.

[Fossum, 1997] Fossum, E. (1997). CMOS image sensors : electronic camera-on-a-chip. *IEEE Transactions on Electron Devices*, 44(10) :1689–1698. Conference Name : IEEE Transactions on Electron Devices.

[Foulkes et al., 2013] Foulkes, A., Rushton, S., and Warren, P. (2013). Flow parsing and heading perception show similar dependence on quality and quantity of optic flow. *Frontiers in Behavioral Neuroscience*, 7.

[Fricker et al., 2022] Fricker, P., Chauhan, T., Hurter, C., and Cottereau, B. (2022). Event-based Extraction of Navigation Features from Unsupervised Learning of Optic Flow Patterns. In *17th International Conference on Computer Vision Theory and Applications*, pages 702–710, Vienne (Online Streaming), France. SCITEPRESS - Science and Technology Publications.

[Fricker et al., 2021] Fricker, P., Chauhan, T., Hurter, C., and Cottereau, B. R. (2021). Modeling the Development of Optic Flow Processing in Primates Using Spiking Neural Networks and Hebbian Learning. Published : Bernstein Conference 2021.

[Gallego et al., 2022] Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A. J., Conradt, J., Daniilidis, K., and Scaramuzza, D. (2022). Event-Based Vision : A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1) :154–180. Conference Name : IEEE Transactions on Pattern Analysis and Machine Intelligence.

[Gallego and Scaramuzza, 2017] Gallego, G. and Scaramuzza, D. (2017). Accurate angular velocity estimation with an event camera. *IEEE Robotics and Automation Letters*, 2(2) :632–639. Publisher : IEEE.

[Ganguli and Simoncelli, 2014] Ganguli, D. and Simoncelli, E. P. (2014). Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural Computation*, 26(10) :2103–2134.

[Gerstner and Kistler, 2002] Gerstner, W. and Kistler, W. M. (2002). *Spiking neuron models : single neurons, populations, plasticity*. Cambridge University Press, Cambridge, U.K. ; New York. OCLC : 57417395.

[Gibson, 1947] Gibson, J. J. (1947). *Motion Picture Testing and Research*. U.S. Government Printing Office. Google-Books-ID : BmUaAAAAIAAJ.

[Gibson, 1950] Gibson, J. J. (1950). *The perception of the visual world*. The perception of the visual world. Houghton Mifflin, Oxford, England. Pages : xii, 242.

[Gibson, 1958] Gibson, J. J. (1958). Visually Controlled Locomotion and Visual Orientation in Animals*. *British Journal of Psychology*, 49(3) :182–194. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2044-8295.1958.tb00656.x.

[Goodale and Milner, 1992] Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1) :20–25.

[Graves et al., 2013] Graves, A., Mohamed, A.-r., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE*

International Conference on Acoustics, Speech and Signal Processing, pages 6645–6649. ISSN : 2379-190X.

[Graziano and Gross, 1994] Graziano, M. and Gross, C. G. (1994). Mapping Space With Neurons. *Current Directions in Psychological Science*, 3(5) :164–167. Publisher : SAGE Publications Inc.

[Grünert and Martin, 2020] Grünert, U. and Martin, P. R. (2020). Cell types and cell circuits in human and non-human primate retina. *Progress in Retinal and Eye Research*, 78 :100844.

[Gu et al., 2010] Gu, Y., Fetsch, C. R., Adeyemo, B., DeAngelis, G. C., and Angelaki, D. E. (2010). Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron*, 66(4) :596–609.

[Gu et al., 2006] Gu, Y., Watkins, P. V., Angelaki, D. E., and DeAngelis, G. C. (2006). Visual and Nonvisual Contributions to Three-Dimensional Heading Selectivity in the Medial Superior Temporal Area. *The Journal of Neuroscience*, 26(1) :73–85.

[Guénot et al., 2022] Guénot, J., Trotter, Y., Fricker, P., Cherubini, M., Soler, V., and Cottereau, B. R. (2022). Optic flow processing in patients with macular degeneration. *Investigative Ophtalmology and Visual Science*.

[Hadjikhani et al., 1998] Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., and Tootell, R. B. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nature Neuroscience*, 1(3) :235–241.

[Hausselt et al., 2007] Hausselt, S. E., Euler, T., Detwiler, P. B., and Denk, W. (2007). A dendrite-autonomous mechanism for direction selectivity in retinal starburst amacrine cells. *PLoS biology*, 5(7) :e185.

[Hebb, 1949] Hebb, D. O. (1949). *The organization of behavior ; a neuropsychological theory*. The organization of behavior ; a neuropsychological theory. Wiley, Oxford, England. Pages : xix, 335.

[Heiberg et al., 2013] Heiberg, T., Kriener, B., Tetzlaff, T., Casti, A., Einevoll, G., and Plesser, H. (2013). Firing-rate models capture essential response dynamics of LGN relay cells. *Journal of computational neuroscience*, 35.

[Herrmann et al., 2011] Herrmann, R., Heflin, S. J., Hammond, T., Lee, B., Wang, J., Gainetdinov, R. R., Caron, M. G., Eggers, E. D., Frishman, L. J., McCall, M. A., and Arshavsky, V. Y. (2011). Rod Vision Is Controlled by Dopamine-Dependent Sensitization of Rod Bipolar Cells by GABA. *Neuron*, 72(1) :101–110. Publisher : Elsevier.

[Hodgkin and Huxley, 1952a] Hodgkin, A. L. and Huxley, A. F. (1952a). The components of membrane conductance in the giant axon of Loligo. *The Journal of Physiology*, 116(4) :473–496. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1952.sp004718.

[Hodgkin and Huxley, 1952b] Hodgkin, A. L. and Huxley, A. F. (1952b). Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo. *The Journal of Physiology*, 116(4) :449–472. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1952.sp004717.

[Hodgkin and Huxley, 1952c] Hodgkin, A. L. and Huxley, A. F. (1952c). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4) :500–544. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1952.sp004764.

[Hodgkin et al., 1952] Hodgkin, A. L., Huxley, A. F., and Katz, B. (1952). Measurement of current-voltage relations in the membrane of the giant axon of Loligo. *The Journal of Physiology*, 116(4) :424–448. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1952.sp004716.

[Hubel and Wiesel, 1962] Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1) :106–154.2.

[Hubel and Wiesel, 1965] Hubel, D. H. and Wiesel, T. N. (1965). RECEPTIVE FIELDS AND FUNCTIONAL ARCHITECTURE IN TWO NONSTRIATE VISUAL AREAS (18 AND 19) OF THE CAT. *Journal of Neurophysiology*, 28 :229–289.

[Hubel and Wiesel, 1968] Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1) :215–243.

[Hubel et al., 1977] Hubel, D. H., Wiesel, T. N., and Stryker, M. P. (1977). Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. *Nature*, 269(5626) :328–330. Number : 5626 Publisher : Nature Publishing Group.

[Hubel et al., 1978] Hubel, D. H., Wiesel, T. N., and Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey. *Journal of Comparative Neurology*, 177(3) :361–379. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.901770302.

[Hübner et al., 2020] Hübner, P., Clintworth, K., Liu, Q., Weinmann, M., and Wursthorn, S. (2020). Evaluation of HoloLens Tracking and Depth Sensing for Indoor Mapping Applications. *Sensors*, 20(4) :1021. Number : 4 Publisher : Multidisciplinary Digital Publishing Institute.

[Ilg, 2008] Ilg, U. J. (2008). The role of areas MT and MST in coding of visual motion underlying the execution of smooth pursuit. *Vision Research*, 48(20) :2062–2069.

[Indiveri et al., 2006] Indiveri, G., Chicca, E., and Douglas, R. (2006). A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity. *IEEE Transactions on Neural Networks*, 17(1) :211–221.

[Jacob et al., 2007] Jacob, V., Brasier, D. J., Erchova, I., Feldman, D., and Shulz, D. E. (2007). Spike timing-dependent synaptic depression in the in vivo barrel cortex of the rat. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 27(6) :1271–1284.

[Jacoby et al., 1996] Jacoby, R., Stafford, D., Kouyama, N., and Marshak, D. (1996). Synaptic inputs to ON parasol ganglion cells in the primate retina. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 16(24) :8041–8056.

[Kandel et al., 2012] Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J., editors (2012). *Principles of Neural Science*. McGraw-Hill Professional Pub, New York, 5th edition edition.

[Kennedy and Bullier, 1985] Kennedy, H. and Bullier, J. (1985). A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey. *Journal of Neuroscience*.

[Kheradpisheh and Masquelier, 2020] Kheradpisheh, S. R. and Masquelier, T. (2020). Temporal Backpropagation for Spiking Neural Networks with One Spike per Neuron. *International Journal of Neural Systems*, 30(06) :2050027. Publisher : World Scientific Publishing Co.

[Kim et al., 1996] Kim, N.-G., Growney, R., and Turvey, M. T. (1996). Optical flow not retinal flow is the basis of wayfinding by foot. *Journal of Experimental Psychology : Human Perception and Performance*, 22(5) :1279–1288.

[Koch et al., 2004] Koch, C., Gazzaniga, M. S., Heatherton, T. F., Ledoux, J. E., and Logothetis, N. (2004). *The Cognitive Neurosciences*. MIT Press.

[Koenderink, 1986] Koenderink, J. J. (1986). Optic flow. *Vision Research*, 26(1) :161–179.

[Kolster et al., 2010] Kolster, H., Peeters, R., and Orban, G. A. (2010). The Retinotopic Organization of the Human Middle Temporal Area MT/V5 and Its Cortical Neighbors. *Journal of Neuroscience*, 30(29) :9801–9820. Publisher : Society for Neuroscience Section : Articles.

[Kuffler, 1953] Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1) :37–68. Publisher : American Physiological Society.

[Lagae et al., 1993] Lagae, L., Raiguel, S., and Orban, G. A. (1993). Speed and direction selectivity of macaque middle temporal neurons. *Journal of Neurophysiology*, 69(1) :19–39.

[Lakshmi et al., 2019] Lakshmi, A., Chakraborty, A., and Thakur, C. S. (2019). Neuromorphic vision : From sensors to event-based algorithms. *WIREs Data Mining and Knowledge Discovery*, 9(4) :e1310. _eprint : https ://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1310.

[Layton and Fajen, 2022] Layton, O. W. and Fajen, B. R. (2022). Distributed encoding of curvilinear self-motion across spiral optic flow patterns. *Scientific Reports*, 12(1) :13393.

[Layton et al., 2012] Layton, O. W., Mingolla, E., and Browning, N. A. (2012). A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects. *Journal of Vision*, 12(1) :20.

[LeCun and Bengio, 1995] LeCun, Y. and Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10) :1995.

[Lee et al., 2020] Lee, C., Kosta, A. K., Zhu, A. Z., Chaney, K., Daniilidis, K., and Roy, K. (2020). Spike-FlowNet : Event-based Optical Flow Estimation with Energy-Efficient Hybrid Neural Networks. *arXiv :2003.06696 [cs]*. arXiv : 2003.06696.

[Lee, 1976] Lee, D. N. (1976). A Theory of Visual Control of Braking Based on Information about Time-to-Collision. *Perception*, 5(4) :437–459. Publisher : SAGE Publications Ltd STM.

[Lee and Aronson, 1974] Lee, D. N. and Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Perception & Psychophysics*, 15(3) :529–532.

[Levitt et al., 1994] Levitt, J. B., Yoshioka, T., and Lund, J. S. (1994). Intrinsic cortical connections in macaque visual area V2 : evidence for interaction between different functional streams. *The Journal of Comparative Neurology*, 342(4) :551–570.

[Li et al., 2018] Li, L., Ni, L., Lappe, M., Niehorster, D. C., and Sun, Q. (2018). No special treatment of independent object motion for heading perception. *Journal of Vision*, 18(4) :19.

[Lichtsteiner et al., 2008] Lichtsteiner, P., Posch, C., and Delbruck, T. (2008). A 128× 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid State Circuits*, 43(2) :566–576. Number : 2 Publisher : IEEE.

[Linsker, 1986] Linsker, R. (1986). From basic network principles to neural architecture : emergence of orientation-selective cells. *Proceedings of the National Academy of Sciences*, 83(21) :8390–8394. Publisher : Proceedings of the National Academy of Sciences.

[Livingstone and Hubel, 1984] Livingstone, M. S. and Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 4(1) :309–356.

[Livingstone and Hubel, 1987] Livingstone, M. S. and Hubel, D. H. (1987). Connections between layer 4B of area 17 and the thick cytochrome oxidase stripes of area 18 in the squirrel monkey. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 7(11) :3371–3377.

[Longuet-Higgins and Prazdny, 1980] Longuet-Higgins, H. C. and Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 208(1173) :385–397. Publisher : Royal Society.

[Lueck et al., 1989] Lueck, C. J., Zeki, S., Friston, K. J., Deiber, M.-P., Cope, P., Cunningham, V. J., Lammertsma, A. A., Kennard, C., and Frackowiak, R. S. J. (1989). The colour centre in the cerebral cortex of man. *Nature*, 340(6232) :386–389. Number : 6232 Publisher : Nature Publishing Group.

[Maass, 1997] Maass, W. (1997). Networks of spiking neurons : The third generation of neural network models. *Neural Networks*, 10(9) :1659–1671.

[Maaten and Hinton, 2008] Maaten, L. v. d. and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86) :2579–2605.

[Markram et al., 1997] Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science (New York, N.Y.)*, 275(5297) :213–215.

[Masquelier and Thorpe, 2010] Masquelier, T. and Thorpe, S. (2010). Learning to recognize objects using waves of spikes and Spike Timing-Dependent Plasticity. *The 2010 International Joint Conference on Neural Networks (IJCNN)*.

[Masquelier and Thorpe, 2007] Masquelier, T. and Thorpe, S. J. (2007). Unsupervised Learning of Visual Features through Spike Timing Dependent Plasticity. *PLOS Computational Biology*, 3(2) :e31. Publisher : Public Library of Science.

[Maunsell and van Essen, 1983] Maunsell, J. H. and van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 3(12) :2563–2586.

[McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4) :115–133.

[MEADOWS, 1974] MEADOWS, J. C. (1974). DISTURBED PERCEPTION OF COLOURS ASSOCIATED WITH LOCALIZED CEREBRAL LESIONS. *Brain*, 97(1) :615–632.

[Meng et al., 2020] Meng, Z., Hu, Y., and Ancey, C. (2020). Using a Data Driven Approach to Predict Waves Generated by Gravity Driven Mass Flows. *Water*, 12(2) :600. Number : 2 Publisher : Multidisciplinary Digital Publishing Institute.

[Mermillod et al., 2010] Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., and Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions : Evidence from neural computation. *Neurocomputing*, 73(13) :2522–2531.

[Mikolov and Zweig, 2012] Mikolov, T. and Zweig, G. (2012). Context dependent recurrent neural network language model. In *2012 IEEE Spoken Language Technology Workshop (SLT)*, pages 234–239.

[Mineault et al., 2012] Mineault, P. J., Khawaja, F. A., Butts, D. A., and Pack, C. C. (2012). Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proceedings of the National Academy of Sciences*, 109(16) :E972–E980. Publisher : Proceedings of the National Academy of Sciences.

[Mink et al., 1981] Mink, J. W., Blumenschine, R. J., and Adams, D. B. (1981). Ratio of central nervous system to body metabolism in vertebrates : its constancy and functional basis. *The American Journal of Physiology*, 241(3) :R203–212.

[Mishkin et al., 1983] Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision : two cortical pathways. *Trends in Neurosciences*, 6 :414–417.

[Movshon and Newsome, 1996] Movshon, J. A. and Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 16(23) :7733–7741.

[Mozafari et al., 2019] Mozafari, M., Ganjtabesh, M., Nowzari-Dalini, A., Thorpe, S. J., and Masquelier, T. (2019). Bio-inspired digit recognition using reward-modulated spike-timing-dependent plasticity in deep convolutional networks. *Pattern Recognition*, 94 :87–95. arXiv : 1804.00227.

[Mueggler et al., 2014] Mueggler, E., Huber, B., and Scaramuzza, D. (2014). Event-based, 6-DOF pose tracking for high-speed maneuvers. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2761–2768. ISSN : 2153-0866.

[Mueggler et al., 2017] Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T., and Scaramuzza, D. (2017). The Event-Camera Dataset and Simulator : Event-based Data for Pose Estimation, Visual Odometry, and SLAM. *The International Journal of Robotics Research*, 36(2) :142–149. arXiv : 1610.08336.

[Nassi and Callaway, 2009] Nassi, J. J. and Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nature Reviews Neuroscience*, 10(5) :360–372. Number : 5 Publisher : Nature Publishing Group.

[Neftci et al., 2019] Neftci, E., Mostafa, H., and Zenke, F. (2019). Surrogate Gradient Learning in Spiking Neural Networks : Bringing the Power of Gradient-Based Optimization to Spiking Neural Networks. *IEEE Signal Processing Magazine*, 36 :51–63.

[Neil and Liu, 2016] Neil, D. and Liu, S.-C. (2016). Effective sensor fusion with event-based sensors and deep network architectures. In *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2282–2285.

[Nguyen et al., 2019] Nguyen, A., Do, T.-T., Caldwell, D. G., and Tsagarakis, N. G. (2019). Real-Time 6DOF Pose Relocalization for Event Cameras With Stacked Spatial LSTM Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[Norman, 2002] Norman, J. (2002). Two visual systems and two theories of perception : An attempt to reconcile the constructivist and ecological approaches. *The Behavioral and Brain Sciences*, 25(1) :73–96 ; discussion 96–144.

[Oja, 1982] Oja, E. (1982). Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3) :267–273.

[Olshausen and Field, 1997] Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set : A strategy employed by V1 ? *Vision Research*, 37(23) :3311–3325.

[Olveczky et al., 2003] Olveczky, B. P., Baccus, S. A., and Meister, M. (2003). Segregation of object and background motion in the retina. *Nature*, 423(6938) :401–408.

[Orban et al., 1986] Orban, G. A., Kennedy, H., and Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey : influence of eccentricity. *Journal of Neurophysiology*, 56(2) :462–480. Publisher : American Physiological Society.

[Orchard et al., 2013] Orchard, G., Benosman, R., Etienne-Cummings, R., and Thakor, N. (2013). *A spiking neural network architecture for visual motion estimation*. Pages : 301.

[Pack and Born, 2001] Pack, C. C. and Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409(6823) :1040–1042.

[Paredes-Vallés et al., 2020] Paredes-Vallés, F., Scheper, K. Y. W., and de Croon, G. C. H. E. (2020). Unsupervised Learning of a Hierarchical Spiking Neural Network for Optical Flow Estimation : From Events to Global Motion Perception. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8) :2051–2064. arXiv : 1807.10936.

[Pearson, 1901] Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11) :559–572. Publisher : Taylor & Francis _eprint : https ://doi.org/10.1080/14786440109462720.

[Pellegrini et al., 2020] Pellegrini, T., Zimmer, R., and Masquelier, T. (2020). Low-activity supervised convolutional spiking neural networks applied to speech commands recognition. *arXiv :2011.06846 [cs, eess]*. arXiv : 2011.06846.

[Peltier et al., 2020] Peltier, N. E., Angelaki, D. E., and DeAngelis, G. C. (2020). Optic flow parsing in the macaque monkey. *Journal of Vision*, 20(10) :8.

[Petkov and Subramanian, 2007] Petkov, N. and Subramanian, E. (2007). Motion detection, noise reduction, texture suppression, and contour enhancement by spatiotemporal Gabor filters with surround inhibition. *Biological Cybernetics*, 97(5-6) :423–439.

[Pfister and Gerstner, 2006] Pfister, J.-P. and Gerstner, W. (2006). Triplets of Spikes in a Model of Spike Timing-Dependent Plasticity. *Journal of Neuroscience*, 26(38) :9673–9682. Publisher : Society for Neuroscience Section : Articles.

[Phinney and Siegel, 1999] Phinney, R. E. and Siegel, R. M. (1999). Stored representations of three-dimensional objects in the absence of two-dimensional cues. *Perception*, 28(6) :725–737.

[Pidoux, 2011] Pidoux, B. (2011). Les Voies et Centres Visuels.

[Pitcher and Ungerleider, 2021] Pitcher, D. and Ungerleider, L. G. (2021). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, 25(2) :100–110.

[Posch et al., 2011] Posch, C., Matolin, D., and Wohlgenannt, R. (2011). A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS. *IEEE Journal of Solid-State Circuits*, 46(1) :259–275. Conference Name : IEEE Journal of Solid-State Circuits.

[Posch et al., 2014] Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B., and Delbruck, T. (2014). Retinomorphic Event-Based Vision Sensors : Bioinspired Cameras With Spiking Output. *Proceedings of the IEEE*, 102(10) :1470–1484. Conference Name : Proceedings of the IEEE.

[Priebe et al., 2003] Priebe, N. J., Cassanello, C. R., and Lisberger, S. G. (2003). The Neural Representation of Speed in Macaque Area MT/V5. *The Journal of Neuroscience*, 23 :5650–5661. Place : US Publisher : Society for Neuroscience.

[Priebe et al., 2006] Priebe, N. J., Lisberger, S. G., and Movshon, J. A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *The Journal of*

Neuroscience : The Official Journal of the Society for Neuroscience, 26(11) :2941–2950.

[Raffi et al., 2013] Raffi, M., Piras, A., Persiani, M., and Squatrito, S. (2013). Importance of optic flow for postural stability of male and female young adults. *European journal of applied physiology*, 114.

[Raffi and Siegel, 2007] Raffi, M. and Siegel, R. M. (2007). A Functional Architecture of Optic Flow in the Inferior Parietal Lobule of the Behaving Monkey. *PLOS ONE*, 2(2) :e200. Publisher : Public Library of Science.

[Raisman, 1969] Raisman, G. (1969). Neuronal plasticity in the septal nuclei of the adult rat. *Brain Research*, 14(1) :25–48.

[Riddell et al., 2019] Riddell, H., Li, L., and Lappe, M. (2019). Heading perception from optic flow in the presence of biological motion. *Journal of Vision*, 19(14) :25.

[Robroek, 2020] Robroek, R. (2020). Creating 3D Interactive Objects from Photographs.

[Rodieck, 1965] Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5(12) :583–601.

[Rogers, 2021] Rogers, B. (2021). Optic Flow : Perceiving and Acting in a 3-D World. *i-Perception*, 12(1) :2041669520987257. Publisher : SAGE Publications.

[Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron : A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65 :386–408. Place : US Publisher : American Psychological Association.

[Rossetti et al., 2017] Rossetti, Y., Pisella, L., and McIntosh, R. D. (2017). Rise and fall of the two visual systems theory. *Annals of Physical and Rehabilitation Medicine*, 60(3) :130–140.

[Royden and Hildreth, 1996] Royden, C. S. and Hildreth, E. C. (1996). Human heading judgments in the presence of moving objects. *Perception & Psychophysics*, 58(6) :836–856.

[Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088) :533–536. Number : 6088 Publisher : Nature Publishing Group.

[Sato et al., 2010] Sato, N., Kishore, S., Page, W. K., and Duffy, C. J. (2010). Cortical neurons combine visual cues about self-movement. *Experimental Brain Research*, 206(3) :283–297.

[Shapley and Enroth-Cugell, 1984] Shapley, R. and Enroth-Cugell, C. (1984). Chapter 9 Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3 :263–346.

[Sheth and Young, 2016] Sheth, B. R. and Young, R. (2016). Two Visual Pathways in Primates Based on Sampling of Space : Exploitation and Exploration of Visual Information. *Frontiers in Integrative Neuroscience*, 10 :37.

[Shipp and Zeki, 1985] Shipp, S. and Zeki, S. (1985). Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature*, 315(6017) :322–324. Number : 6017 Publisher : Nature Publishing Group.

[Sjöström et al., 2001] Sjöström, P. J., Turrigiano, G. G., and Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32(6) :1149–1164.

[Snowden et al., 1992] Snowden, R., Treue, S., and Andersen, R. (1992). The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, 88 :389–400.

[Son et al., 2017] Son, B., Suh, Y., Kim, S., Jung, H., Kim, J.-S., Shin, C., Park, K., Lee, K., Park, J., Woo, J., Roh, Y., Lee, H., Wang, Y., Ovsiannikov, I., and Ryu, H. (2017). 4.1 A 640×480 dynamic vision sensor with a 9μm pixel and 300Meps address-event representation. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 66–67. ISSN : 2376-8606.

[Steffen et al., 2019] Steffen, L., Reichard, D., Weinland, J., Kaiser, J., Roennau, A., and Dillmann, R. (2019). Neuromorphic Stereo Vision : A Survey of Bio-Inspired Sensors and Algorithms. *Frontiers in Neurorobotics*, 13.

[Steinmetz et al., 2022] Steinmetz, S. T., Layton, O. W., Powell, N. V., and Fajen, B. R. (2022). A Dynamic Efficient Sensory Encoding Approach to Adaptive Tuning in Neural Models of Optic Flow Processing. *Frontiers in Computational Neuroscience*, 16.

[Stoffregen, 1985] Stoffregen, T. A. (1985). Flow structure versus retinal location in the optical control of stance. *Journal of Experimental Psychology : Human Perception and Performance*, 11(5) :554–565. Place : US Publisher : American Psychological Association.

[Stone, 2012] Stone, J. V. (2012). *Vision and Brain : How We Perceive the World*. MIT Press, Cambridge, MA, USA.

[Stromatias et al., 2017] Stromatias, E., Soto, M., Serrano-Gotarredona, T., and Linares-Barranco, B. (2017). An event-driven classifier for spiking neural networks fed with synthetic or dynamic vision sensor data. *Frontiers in neuroscience*, 11 :350. Publisher : Frontiers.

[Takahashi et al., 2007] Takahashi, K., Gu, Y., May, P. J., Newlands, S. D., DeAngelis, G. C., and Angelaki, D. E. (2007). Multimodal Coding of Three-Dimensional Rotation and Translation in Area MSTd : Comparison of Visual and Vestibular Selectivity. *The Journal of Neuroscience*, 27(36) :9742–9756.

[Tanaka and Farah, 1993] Tanaka, J. W. and Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology A : Human Experimental Psychology*, 46A :225–245. Place : United Kingdom Publisher : Taylor & Francis.

[Tcheang et al., 2005] Tcheang, L., Gilson, S. J., and Glennerster, A. (2005). Systematic distortions of perceptual stability investigated using immersive virtual reality. *Vision Research*, 45(16) :2177–2189.

[Thorpe et al., 2001] Thorpe, S., Delorme, A., and Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks : The Official Journal of the International Neural Network Society*, 14(6-7) :715–725.

[Todd, 1995] Todd, J. T. (1995). The visual perception of three-dimensional structure from motion. In *Perception of space and motion*, Handbook of perception and cognition (2nd ed.), pages 201–226. Academic Press, San Diego, CA, US.

[Tootell et al., 1997] Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., Sereno, M. I., and Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 17(18) :7060–7078.

[Tortora and Derrickson, 2016] Tortora, G. J. and Derrickson, B. (2016). *Eléments d'anatomie et de physiologie*.

[Ungerleider and Mishkin, 1982] Ungerleider, L. G. and Mishkin, M. (1982). *Two cortical visual systems*. Analysis of visual behavior. MIT Press, Cambridge, Mass. Meeting Name : NATO Advanced Study Institute.

[van den Berg, 1992] van den Berg, A. V. (1992). Robustness of perception of heading from optic flow. *Vision Research*, 32(7) :1285–1296.

[Van Essen et al., 2001] Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B., Bakircioglu, M., and Miller, M. I. (2001). Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Research*, 41(10-11) :1359–1378.

[Van Essen and Maunsell, 1983] Van Essen, D. C. and Maunsell, J. H. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, 6 :370–375. Place : Netherlands Publisher : Elsevier Science.

[Van Essen et al., 1981] Van Essen, D. C., Maunsell, J. H. R., and Bixby, J. L. (1981). The middle temporal visual area in the macaque : Myeloarchitecture, connections, functional properties and topographic organization. *Journal of Comparative Neurology*, 199(3) :293–326. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.901990302.

[VanRullen et al., 2005] VanRullen, R., Guyonneau, R., and Thorpe, S. J. (2005). Spike times make sense. *Trends in Neurosciences*, 28(1) :1–4.

[Verweij et al., 2003] Verweij, J., Hornstein, E. P., and Schnapf, J. L. (2003). Surround Antagonism in Macaque Cone Photoreceptors. *Journal of Neuroscience*, 23(32) :10249–10257. Publisher : Society for Neuroscience Section : Behavioral/Systems/Cognitive.

[Warren and Rushton, 2007] Warren, P. A. and Rushton, S. K. (2007). Perception of object trajectory : Parsing retinal motion into self and object movement components. *Journal of Vision*, 7(11) :2.

[Warren and Rushton, 2008] Warren, P. A. and Rushton, S. K. (2008). Evidence for flow-parsing in radial flow displays. *Vision Research*, 48(5) :655–663.

[Warren and Rushton, 2009] Warren, P. A. and Rushton, S. K. (2009). Optic Flow Processing for the Assessment of Object Movement during Ego Movement. *Current Biology*, 19(18) :1555–1560.

[Warren, 1976] Warren, R. (1976). The perception of egomotion. *Journal of Experimental Psychology : Human Perception and Performance*, 2(3) :448–456. Place : US Publisher : American Psychological Association.

[Warren et al., 2001] Warren, W. H., Kay, B. A., Zosh, W. D., Duchon, A. P., and Sahuc, S. (2001). Optic flow is used to control human walking. *Nature Neuroscience*, 4(2) :213–216. Number : 2 Publisher : Nature Publishing Group.

[Warren et al., 1988] Warren, W. H., Morris, M. W., and Kalish, M. (1988). Perception of translational heading from optical flow. *Journal of Experi-*

mental Psychology : Human Perception and Performance, 14(4) :646–660. Place : US Publisher : American Psychological Association.

[Warren and Saunders, 1995] Warren, W. H. and Saunders, J. A. (1995). Perceiving Heading in the Presence of Moving Objects. *Perception*, 24(3) :315–331. Publisher : SAGE Publications Ltd STM.

[Woźniak et al., 2020] Woźniak, S., Pantazi, A., Bohnstingl, T., and Eleftheriou, E. (2020). Deep learning incorporating biologically inspired neural dynamics and in-memory computing. *Nature Machine Intelligence*, 2(6) :325–336. Number : 6 Publisher : Nature Publishing Group.

[Wässle, 2004] Wässle, H. (2004). Parallel processing in the mammalian retina. *Nature Reviews Neuroscience*, 5(10) :747–757. Number : 10 Publisher : Nature Publishing Group.

[Young et al., 2007] Young, J. M., Waleszczyk, W. J., Wang, C., Calford, M. B., Dreher, B., and Obermayer, K. (2007). Cortical reorganization consistent with spike timing-but not correlation-dependent plasticity. *Nature Neuroscience*, 10(7) :887–895.

[Zeki, 2003] Zeki, S. (2003). Improbable areas in the visual brain. *Trends in Neurosciences*, 26(1) :23–26. Publisher : Elsevier.

[Zeki, 1978] Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, 274(5670) :423–428. Number : 5670 Publisher : Nature Publishing Group.

[Zenke et al., 2021] Zenke, F., Bohté, S. M., Clopath, C., Comşa, I. M., Göltz, J., Maass, W., Masquelier, T., Naud, R., Neftci, E. O., Petrovici, M. A., Scherr, F., and Goodman, D. F. M. (2021). Visualizing a joint future of neuroscience and neuromorphic engineering. *Neuron*, 109(4) :571–575.

[Zhang et al., 1998] Zhang, L. I., Tao, H. W., Holt, C. E., Harris, W. A., and Poo, M. (1998). A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395(6697) :37–44.

[Zhu et al., 2018a] Zhu, A. Z., Thakur, D., Özaslan, T., Pfrommer, B., Kumar, V., and Daniilidis, K. (2018a). The Multivehicle Stereo Event Camera Dataset : An Event Camera Dataset for 3D Perception. *IEEE Robotics and Automation Letters*, 3(3) :2032–2039. Conference Name : IEEE Robotics and Automation Letters.

[Zhu et al., 2018b] Zhu, A. Z., Yuan, L., Chaney, K., and Daniilidis, K. (2018b). EV-FlowNet : Self-Supervised Optical Flow Estimation for Event-based Cameras. *Robotics : Science and Systems XIV*. arXiv : 1802.06898.

[Zhu et al., 2019] Zhu, A. Z., Yuan, L., Chaney, K., and Daniilidis, K. (2019). Unsupervised event-based learning of optical flow, depth, and egomotion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 989–997.

Annexe A

NeuroSoc par Yumain

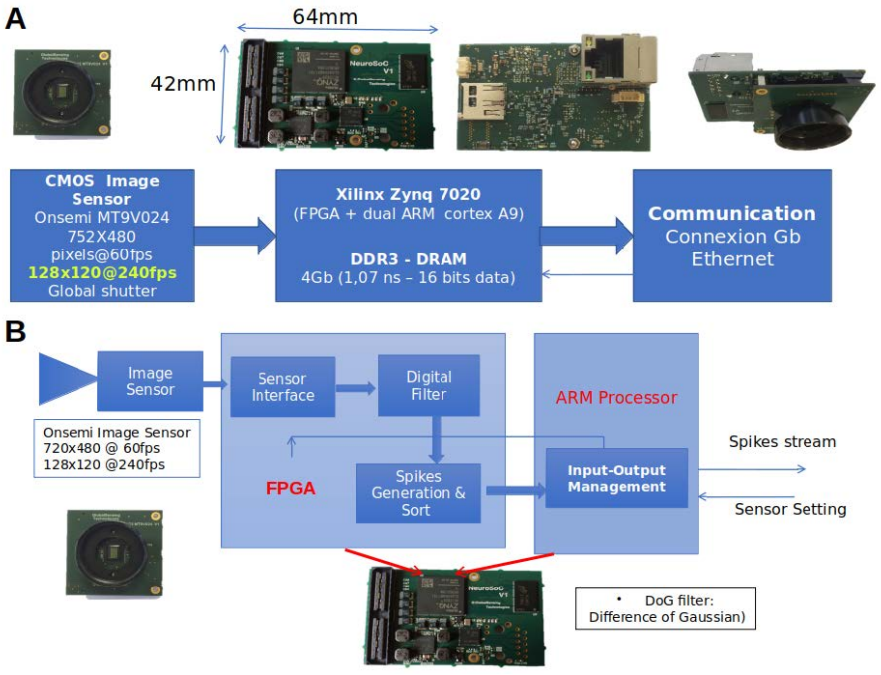


FIGURE A.1 – Composants et chaîne de traitement de la plateforme NeuroSoc associée à son capteur d'images CMOS tiré de [Debat et al., 2021, Debat, 2021]. A) Le capteur CMOS associé à la plateforme NeuroSoc présentant son SoC Zynq, ses blocs de mémoire vive et de gestion de communication par protocole ethernet. B) La chaîne de traitement vidéo et de génération d'événements effectués par le SoC Zynq.

Les filtres des différents modèles de caméras événementielles n'utilisent que l'information temporelle en se basant sur la variation de luminance au sein de la scène capturée. En n'utilisant que cette information temporelle, ces caméras perdent la composante spatiale du filtrage du système visuel le rendant spatio-temporel. La plateforme NeuroSoc développée par *Yumain* propose en plus d'une différence temporelle des pixels de la scène capturée, un filtrage spatial par DoG ou filtres de Gabor orientés. La génération des événements se fait alors à l'aide d'un capteur synchrone, auquel on rajoute les étages de filtrages spatiaux et temporels, visant à extraire les propriétés des événements générés : position, latence et polarité.

Couplée à un capteur CMOS, la plateforme NeuroSoc devient alors grâce à ce capteur d'images le *Spike Event Sensor*, capteur permettant la capture d'images, leur filtrage et la génération des événements correspondants. Sa description suivante est permise grâce au travail de thèse au CerCo de Guillaume Debat [Debat et al., 2021, Debat, 2021].

Le capteur d'images utilisé est un capteur CMOS MT9V024 développé par *ON Semiconductor* avec un mode d'acquisition en *global-shutter* à une fréquence de 60 images par seconde pour une résolution de 752×480 pixels et allant jusqu'à une fréquence d'acquisition de 240 images par seconde pour une résolution de 128×120 pixels. Le mode *global-shutter* correspond à un mode d'acquisition instantanée de l'image garantisant une fréquence d'acquisition élevée grâce à un faible temps d'exposition. Sa carte électronique présente essentiellement le *System on Chip* (SoC) Zynd 7020 de *Xilinx* et d'un bloc mémoire vive de 4 gigabits. Le SoC Zynq 7020 se décompose en plusieurs parties, un FPGA et un processeur double coeur ARM Cortex-A9.

Le traitement des images entrant consituant le signal vidéo capturé par le capteur CMOS se voit être transmis au SoC Zynq chargé d'effectuer le filtrage spatio-temporel. Le premier filtrage opéré, à l'instar des caméras événementielles, est un filtrage temporel. Il est effectué en soustrayant l'image capturée à l'instant t à l'image capturée précédemment à l'instant $t-1$. Les différents composants du capteur *Spike Event Sensor* et sa chaîne de traitement sont illustrés par la figure A.1. Une fois ce filtrage effectué, l'image obtenue est filtrée spatialement par un noyau DoG de dimension

5×5 ayant pour valeur centrale 16 et normalisé par les valeurs -1 en périphérie précisé dans la description matérielle du FPGA et pouvant être changé en taille et en nature. Ce traitement permet finalement d'obtenir les événements de la scène capturée, ces derniers résultants des valeurs d'intensités positives et négatives obtenus après les filtrages des images et correspondant alors aux événements polarisés de type ON et OFF. Les valeurs d'intensités obtenues sont alors seuillées selon leur valeur absolue, ne conservant alors que les valeurs d'intensités dépassant la valeur de seuil définie, reduisant ainsi le bruit potentiellement généré, et déterminé manuellement selon les conditions d'acquisition et son contexte. Afin d'obtenir la latence de chaque événement généré, cette dernière se retrouve étant inversement proportionnelle à leurs valeurs d'intensité [Thorpe et al., 2001, VanRullen et al., 2005, Masquelier and Thorpe, 2007, Chauhan et al., 2018].

Les valeurs ainsi obtenues de position, latence et polarité de chaque événements sont ensuite transmises au processeur ARM. Celui se charge de l'envoi des événements via le port ethernet de la plateforme NeuroSoc. Il permet également d'intervenir sur les paramètres de seuillage ou de captation vidéo. Les événements ainsi générés et récupérés sont alors exploitables et interprétables après un filtrage spatio-temporel et un seuillage garantissant un rapport signal sur bruit et une pertinence des données les plus élevés possible.