



Cyril MONIER
CYRIL.MONIER@EPFL.CH
LABORATOIRE DE NEUROSCIENCE DES MICROCIRCUITS,
FACULTÉ SV, EPFL

DE L'APPROCHE *BOÎTE NOIRE* AU CONNEXIONNISME, CONFRONTATION DES DIFFÉRENTES MÉTHODES D'INVESTIGATION DU SYSTÈME VISUEL

P our comprendre comment marche une machine, il ne suffit pas d'énumérer ses parties ou de définir sa fonction, il faut décrire les interrelations entre ses composants, lesquelles définissent les transformations que peut accomplir cette machine. Ainsi, si l'on veut reproduire une machine, on ne tiendra compte que des propriétés qui satisfont aux interrelations voulues, conduisant à la séquence de transitions recherchée. Nous définirons que l'ensemble des relations qui définissent une machine comme une unité constitue son organisation, l'ensemble des relations effectives entre les composants présents sur une machine concrète dans un espace donné constitue sa structure et enfin l'ensemble des opérations effectuées entre ce qui est défini comme les entrées et les sorties de la machine décrit les fonctions que l'on va attribuer à celle-ci.

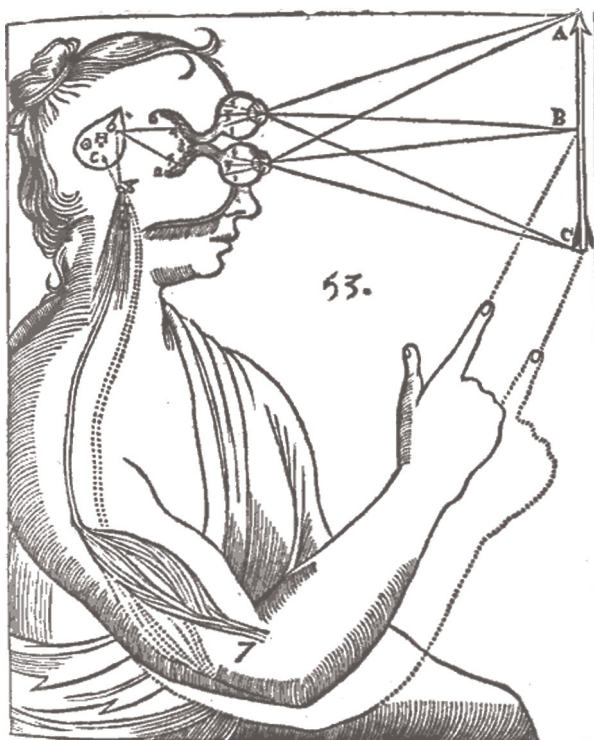
Les systèmes vivants peuvent être assimilés à des machines. Une unité vivante peut ainsi être définie par son organisation, indépendamment de sa structure, de la matérialité au sein de laquelle cette organisation est incorporée. Mais toute explication complète d'un système biologique prendra en compte son organisation, sa structure, comme un exemple de cette organisation, et enfin sera décrit d'un point de vue fonctionnel. Nous allons illustrer à présent les principaux outils théoriques et pratiques qui ont été développés pour étudier ces différents aspects en prenant comme exemple l'étude du système visuel des mammifères. Sommairement, d'un point de vue anatomique, celui-ci se compose dans sa première partie d'un capteur sensoriel, la rétine, connecté à un noyau thalamique, le corps genouillé latéral qui est interconnecté de manière réciproque avec le cortex visuel primaire, lui-même interconnecté avec les autres aires corticales visuelles supérieures.

Toutes les études expérimentales utilisent un cadre théorique plus ou moins explicite aux yeux de l'expérimentateur. Il est pourtant essentiel de maîtriser parfaitement celui-ci afin d'en comprendre les limitations et les pièges. Nous allons présenter ici les deux cadres théoriques, le plus classique pour le système visuel étant la théorie des systèmes (linéaires et non-linéaires) et le plus récent étant l'application de la théorie des systèmes dynamiques à l'étude des réseaux de neurones. Au-delà du cadre théorique, ce sont deux conceptions différentes qui s'opposent ici, celle qui fait du système visuel une simple machine à transformer les signaux entrants en signaux sortants, et celle qui donne au système sa cohérence interne propre. Nous essayerons pourtant de montrer la complémentarité possible de ces approches au niveau expérimental.

Selon la théorie des systèmes, un système est une collection d'éléments qui sont en interaction et dont l'ensemble produit un traitement. En général, un système est défini comme recevant un signal d'entrée de l'extérieur (*input*) et générant un signal en sortie (*output*). La relation entre l'entrée et la sortie d'un système représente une transformation du signal et caractérise la fonction du système. Ainsi la théorie des systèmes produit un cadre d'étude possible pour l'identification fonctionnelle d'un système physique à travers l'examen de la relation entrée-sortie, cette démarche étant classiquement appelée l'approche **boîte noire**. Pour l'étude fonctionnelle du système visuel l'entrée sera le stimulus visuel projeté sur la rétine et la sortie le nombre d'impulsions (les potentiels d'action) d'une cellule nerveuse (un neurone) située par exemple dans le cortex visuel primaire. Les potentiels d'action sont des événements électriques de type tout ou rien qui sont émis par un neurone via son axone; ils permettent soit d'exciter, soit d'inhiber les neurones connectés par celui-ci. Ce qui nous intéresse ici, c'est le nombre de potentiels d'action émis durant un certain temps, c'est-à-dire la fréquence de décharge du neurone. Au cours d'une expérience, les potentiels d'action sont enregistrés à l'aide de fines électrodes placées à côté du neurone pour les électrodes dites *extracellulaires* et dans le neurone pour les électrodes dites

intracellulaires. Les différences de potentiels à la pointe de ces électrodes sont de l'ordre de quelques centaines de microvolts pour les potentiels extracellulaires et de quelques dizaines de millivolts pour les potentiels intracellulaires. Ces différences de potentiels sont amplifiées par des amplificateurs opérationnels et enregistrées à chaque instant dans une mémoire digitale à l'aide d'une carte d'acquisition convertissant les signaux analogiques en signaux digitaux.

Ainsi le but de la théorie des systèmes appliquée aux neurones du système visuel est de caractériser l'opération réalisée entre le stimulus (l'entrée) et la décharge (la sortie), cette fonction sera considérée comme le champ récepteur du neurone. Le terme de champ récepteur, introduit par Sherrington au début du siècle, a été appliqué en 1938 par Hartline aux neurones du système visuel comme une surface rétinienne qui, stimulée par un signal lumineux, module l'activité d'un neurone. C'est à partir des années soixante, avec les célèbres travaux de Hubel et Wiesel, que la notion de champ récepteur est devenue une construction centrale dans le cadre de l'étude analytique du système visuel. Ils ont montré, en particulier, que les neurones du cortex visuel primaire répondent de manière sélective à l'orientation et à la direction d'une barre en mouvement dans une certaine position de l'espace.



LE SYSTÈME VISUEL VU PAR DESCARTES
VISION – RENÉ DESCARTES – 1644

Dans ces premières études, les stimuli utilisés étaient des points lumineux ou des barres, les neurones du cortex visuel répondant bien à ce type de stimuli. Actuellement, des stimuli plus complexes ont été élaborés afin de caractériser directement le champ récepteur du neurone avec le formalisme de la théorie des systèmes. Celui-ci sera considéré dans un premier temps comme linéaire, ce type de systèmes étant relativement simple à étudier puisqu'il est entièrement caractérisé dans le domaine spatio-temporel par sa réponse impulsionnelle. Celle-ci peut être obtenue en mesurant la réponse à un stimulus impulsionnel en chaque position de l'espace. Ainsi, une méthode permettant de définir le champ récepteur dans le domaine spatio-temporel consiste à utiliser une séquence pseudo-aléatoire de stimuli élémentaires, proche de l'impulsion, présentés de manière transitoire et sans pause, en différents points de l'espace. Ces stimuli sont appelés **bruit blanc** car ils contiennent en première approximation toutes les fréquences (son spectre de fréquence est plat, le terme blanc est un analogue de la composition spectrale de la lumière blanche). La fonction impulsionnelle est obtenue simplement par corrélation croisée entre la séquence de stimulation et la décharge du neurone.

Une relation d'entrée-sortie d'un système linéaire peut être également décrite dans le domaine fréquentiel par sa fonction de transfert. Par exemple les stimulations par des réseaux de modulation sinusoïdale de luminance vont permettre de définir les champs récepteurs dans le domaine fréquentiel. En effet, un stimulus périodique peut, quelle que soit sa complexité, être décomposé en différentes sinusoïdes de fréquences, d'amplitude, et de phases variées (décomposition par une série de Fourier). Il est alors nécessaire d'utiliser un grand nombre de stimuli sinusoïdaux avec différents paramètres pour caractériser la fonction de transfert complète (en explorant tout le spectre utile des fréquences spatiales et temporelles).

Si l'opération réalisée par un neurone du système visuel est linéaire, sa fonction de transfert (ou sa réponse impulsionnelle) ainsi obtenue va permettre de prédire la réponse à n'importe quel stimulus par simple convolution entre ce stimulus et cette fonction. Celle-ci pourra rendre compte alors de l'ensemble des propriétés fonctionnelles, comme la sélectivité à l'orientation, à la direction, à la fréquence temporelle ou encore à la fréquence spatiale. Par contre, si la fonction de transfert est non-linéaire, chaque stimulus va mettre en jeu des propriétés particulières dont il faudra tenir compte pour pouvoir prédire la réponse.

Les études réalisées pour la sélectivité à la fréquence spatiale, la sélectivité à l'orientation et la sélectivité à la direction montrent que les champs récepteurs des neurones du cortex visuel primaire présentent en réalité un certain nombre de non-linéarités. Au moins deux cas de non-linéarités peuvent être considérés: une non-linéarité dite **intensive** ou **statique** qui porte directement sur la composante linéaire, que l'on peut estimer en utilisant un seul stimulus, et une non-linéarité provenant des interactions entre plusieurs stimuli, qui ne peut être mesurée qu'en utilisant au moins deux stimuli simultanés (pour une non-linéarité spatiale) ou en décalage dans le temps (pour une non-linéarité temporelle). L'analyse des systèmes non-linéaires repose sur ce formalisme, à une composante linéaire s'ajoutant des composantes non-linéaires d'interaction. Dans cet esprit, Wiener a proposé en 1958 une description mathématique d'une relation d'entrée-sortie pour des systèmes non-linéaires. La réponse à un stimulus bruit blanc est décomposée en série de fonctions, nommées noyaux G de Wiener, dont chaque fonction définit un ordre d'interaction non-linéaire. Le noyau d'ordre 1 est équivalent à la relation linéaire décrite précédemment et la relation d'entrée-sortie va être d'autant plus précise (par rapport à l'erreur d'ajustement) que l'on va augmenter le nombre de termes d'ordre supérieur considérés.

Une fois que le système a été ainsi caractérisé, l'étape suivante consiste à identifier la structure du système, structure étant pris ici dans le sens de la théorie des systèmes, c'est-à-dire dans le sens d'une organisation par rapport à notre définition précédente. Une fois les estimations des noyaux réalisées, il est possible de classer ces systèmes en plusieurs classes structurales. Par exemple le modèle de Wiener (L-N) a une structure sérielle composée d'un système linéaire suivi par un système non-linéaire. Une fois la structure du système ainsi mise en place, il reste à estimer la valeur des paramètres pour les éléments du système. Ces différentes étapes étant réalisées, il devient possible de définir un modèle d'étude pouvant générer des prédictions qui seront testées ensuite expérimentalement.

Une des ambitions avouées de cette approche est que l'organisation de ces systèmes ainsi définie va correspondre plus ou moins à l'organisation anatomo-fonctionnelle du système biologique. Ainsi par exemple deux principaux types de champs récepteurs (CRs) ont été décrits dès les années soixante par Hubel et Wiesel dans le cortex visuel primaire, les CRs dits **Simple** et **Complexes**. Les CRs Simple se caractérisent par leur propriété de linéarité et les CRs complexes par leur non-linéarité. Ainsi l'observation que les champs récepteurs des cellules Simple étaient en première approximation linéaires et ceux des cellules Complexes non-linéaire, a conduit Hubel et Wiesel à proposer un modèle dit *feed-forward*, c'est-à-dire sériel et en boucle ouverte de l'organisation du cortex visuel primaire. L'essence de ce modèle est que le champ récepteur d'une cellule Simple est constitué d'une première étape de sommation linéaire dans laquelle les entrées pré-synaptiques thalamiques sont sommées, suivie d'une étape non-linéaire de rectification dans laquelle le seuil de décharge des potentiels d'action filtre les faibles réponses évoquées par les stimuli non spécifiques. Ce modèle est hiérarchique, les cellules Simple sont considérées comme le premier niveau de traitement cortical et les cellules Complexes représenteraient le deuxième niveau hiérarchique de traitement cortical et ne recevraient leurs entrées des cellules Simple.

L'apparente simplicité de ce modèle *feed-forward* semble néanmoins difficile à réconcilier avec la diversité et le nombre d'entrées synaptiques intracorticales arrivant sur une cellule. Anatomiquement la majorité des contacts synaptiques excitateurs sur les cellules corticales sont produits par d'autres neurones corticaux, également dans la couche qui reçoit

les afférents thalamiques. De même 80% des connexions entre le thalamus et le cortex se font dans le sens cortico-thalamique, plutôt que dans le sens *feed-forward* thalamo-cortical. Ces quelques données anatomiques (et de nombreuses évidences électrophysiologiques) mettent en évidence que le système n'est pas organisé de manière sérielle mais plutôt avec de nombreux bouclages en retour.

Une possibilité serait d'ajuster le modèle *feed-forward* en incorporant petit à petit ces connexions récurrentes, mais le fait de conserver le formalisme de la théorie des systèmes privilégiant les schémas sériels conduit rapidement à une impasse, les propriétés dynamiques des circuits récurrents n'étant pas prises en compte. Pour étudier celles-ci, il est nécessaire d'effectuer un changement de perspective. Sortir de la conception que le cerveau peut être étudié comme un simple transducteur convertissant des messages d'entrée en messages de sortie et traiter celui-ci comme un système dynamique *autonome*, c'est-à-dire clos au niveau de son information et de son organisation, sans entrée ni sortie, étant lui-même la source de ses déterminations. L'accent est mis ainsi sur la cohérence interne et l'autonomie du cerveau vu comme un réseau complexe de calculateur élémentaire en interaction. Au cœur même du néoconnexionnisme, terme large caractérisant ces différentes approches, le courant *Attractor Neural Network* animé principalement par des physiciens étudie les propriétés émergentes de réseaux de neurones fortement interconnectés et s'intéresse non plus seulement à leurs capacités computationnelles, mais à leurs comportements propres encore appelés attracteurs, terme emprunté à la théorie des systèmes dynamiques. De tels réseaux possèdent une multiplicité d'attracteurs et ils tendent vers l'un ou l'autre d'entre eux en fonction de ces conditions initiales. La *vie* d'un réseau peut ainsi se concevoir comme une trajectoire dans son paysage d'attracteurs, le passage de l'un à l'autre résultant de perturbations ou de chocs en provenance du monde extérieur. A noter qu'il n'y a alors plus de corrélation entre les opérations réalisées par le système et l'organisation anatomo-fonctionnelle de celui-ci car c'est le fonctionnement global du réseau qui va faire émerger cette opération et non pas des modules opérationnels mis bout à bout.

Comme nous l'avons énoncé en tout début, ce sont deux conceptions différentes qui s'opposent ici, celle qui fait du système visuel une simple machine à transformer les signaux entrants en signaux sortants, et celle qui donne au système sa cohérence interne propre, dont les caractéristiques ne sont pas le reflet des stimulations sensorielles. Néanmoins ces deux conceptions vont être complémentaires dans l'étude expérimentale du système visuel, le but final étant d'obtenir l'organisation du système et les lois régissant sa dynamique.

Ainsi, un expérimentateur va produire de nombreuses données quantitatives structurelles sur l'anatomie des différents neurones, leurs propriétés physiologiques intrinsèques et leurs modes d'interaction au niveau synaptique, permettant de définir la connectivité du réseau et les propriétés dynamiques de celui-ci. D'un autre côté il va décrire le réseau neuronal comme un système produisant une opération entre une entrée et une sortie qu'il aura lui-même définie et en corrélant l'aspect structurel et fonctionnel, il va faire émerger une organisation. L'outil informatique va ensuite permettre de simuler et tester différentes organisations en produisant des modèles de réseaux de neurones. Il faut vérifier ensuite si le réseau ainsi simulé, va reproduire l'opération entre l'entrée et la sortie et d'autre part si les paramètres utilisés pour le simuler sont proches des paramètres mesurés sur la structure biologique.

La caractérisation de l'opération entre l'entrée et la sortie est utilisée ici comme une description fine du fonctionnement du système et non plus comme un moyen d'accéder directement à son organisation. ■

