
TRAITEMENT VISUEL RAPIDE DE SCÈNES NATURELLES CHEZ LE SINGE, L'HOMME ET LA MACHINE : UNE VISION QUI VA DE L'AVANT...

Arnaud DELORME

CERCO, Centre de Recherche Cerveau et Cognition
133 route de Narbonne, F-31062 TOULOUSE Cedex, France
Tél : +33 5 62 17 28 39 — Fax : +33 5 62 17 28 09

Mel : arno@cerco.ups-tlse.fr

Toile : <http://www.cerco.ups-tlse.fr/~arno/> — <http://si8000.ups-tlse.fr/arno/>

L'énigme du cerveau est l'un des défis les plus ambitieux lancés à la science qui, aujourd'hui, possède enfin des outils d'investigation performants pour en résoudre certains mystères. Dès le milieu du 20^{ème} siècle, même si cet héritage est aujourd'hui renié, l'approche cybernétique a posé les bases d'une étude scientifique des processus cognitifs. En continuité avec ces travaux, à la frontière entre neurosciences et intelligence artificielle, les neurosciences computationnelles tentent de comprendre les formidables capacités de calcul du cerveau, notamment l'efficacité du traitement d'images par le système visuel. Un travail de thèse avec un tel titre se doit de commencer sur une note d'humilité et même s'il est possible que mon approche soit novatrice, je tiens à souligner que le terme "qui va de l'avant" fait également référence à la propagation de l'information visuelle *en avant* à laquelle je m'intéresse tout particulièrement.

Au sein de l'organisation cérébrale, le système visuel est très intéressant car, du fait de son organisation hiérarchique, il établit un lien fort entre des représentations à différents niveaux. On peut aborder l'étude du système visuel au niveau de neurones individuels ou de populations de neurones - allant de quelques milliers de neurones en champs de potentiels à quelques millions en imagerie cérébrale - et tenter d'établir un lien entre le comportement des sujets et l'activité de ces neurones. Il est également possible en psychologie expérimentale d'analyser les réponses de sujets à des stimuli visuels et d'inférer les processus neuronaux sous-jacents. Des modèles permettent ensuite de vérifier ces hypothèses et d'établir un pont entre le comportement des neurones et celui des individus. Ces approches combinées, à différents niveaux, seront sans aucun doute nécessaires pour mettre à jour les processus à l'origine de la formidable efficacité du système visuel et c'est dans ce cadre que se place mon travail de thèse. M'intéressant plus aux populations de neurones qu'aux comportements des neurones individuels, j'ai utilisé une approche de psychologie expérimentale et basé mon travail de modélisation à la fois sur les résultats que j'ai obtenus en psychophysique et sur les contraintes de structure et de connectivité au sein du système visuel.

Mon travail de thèse est donc un double travail expérimental et de modélisation. Dans la partie expérimentale, j'ai tenté d'appréhender les processus mis en jeu dans le système visuel pour la catégorisation ultra-rapide d'images naturelles brièvement flashées. Je m'intéresse en particulier aux facteurs clefs intervenant dans la précision et dans la rapidité des processus visuels. On présente brièvement (20-30 ms) des photographies contenant ou non des animaux au sujet qui doit relâcher un bouton quand l'image contient un animal. Le paradigme expérimental de la tâche go-nogo permet d'étudier le traitement d'une catégorie abstraite - dans nos expériences, la catégorie "animal" - dans des conditions bien contrôlées. La tâche, bien qu'extrêmement simple, fait appel à des traitements très complexes dans le système visuel et l'on connaît encore très mal la façon dont ils s'effectuent. Mes études ont porté sur les processus de catégorisation chez les sujets humains mais aussi chez les singes macaques. Le singe macaque réalise cette tâche avec une précision légèrement inférieure à celle de l'homme mais avec une plus grande rapidité. Les différences et les similitudes entre les deux espèces peuvent en effet nous renseigner sur les différentes stratégies utilisées par les systèmes visuels des primates. Mon approche consiste à contraindre les conditions dans lesquelles est effectuée la catégorisation pour déterminer le rôle à la fois des propriétés intrinsèques des images - couleur, luminance, nombre d'animaux présents, parties visibles de leurs corps, espèce de l'animal... - mais aussi de leurs propriétés extrinsèques - condition de présentation, effet de séquence, familiarité du stimulus, consigne... L'analyse de l'influence de ces caractéristiques sur la catégorisation me permet d'inférer, dans une certaine mesure, la dynamique de traitement des images par le système visuel. Les implications pour la catégorisation ultra-rapide sont de plusieurs ordres. Nos résultats suggèrent que le traitement des images par le système visuel pourrait s'effectuer de façon relativement automatique et précablée du fait de l'absence d'effet sur les temps de réaction précoces de la familiarité de l'image et de l'image précédant l'image catégorisée. Cette hypothèse est renforcée par les résultats d'une expérience où l'on simplifie la tâche et où les sujets doivent catégoriser une image unique parmi des distracteurs variés : le fait que les aires corticales activées, localisées à partir des potentiels évoqués, soient similaires dans la tâche de catégorisation animal/non animal et dans la tâche de détection de cible unique implique un traitement visuel relativement automatique. De plus la catégorisation rapide semble être fondée sur la forme des cibles pour au moins trois raisons. Premièrement, la catégorisation rapide ne dépend pas des informations chromatiques, l'absence de couleur dans les photographies n'influençant pas les temps de réaction précoces des sujets. Deuxièmement, certains éléments diagnostiques de formes précises (comme la présence d'une bouche ou de pattes) influencent la rapidité des réponses et laissent donc penser que leur présence

intervient fortement dans le traitement visuel. Enfin, la catégorisation ne dépendrait pas des attributs sémantiques des images comme le montre l'absence de différence entre les performances des sujets sur des images d'animaux communs (dont on connaît le nom) et d'animaux rares. Les contraintes sur la vitesse de catégorisation sont également très importantes. Pour le singe notamment, on obtient des réponses spécifiques sur les cibles dès 180 ms. Dans ce court laps de temps, l'information sur le stimulus doit transiter par toutes les aires visuelles, par les aires frontales et motrices pour finalement produire une réponse comportementale. En considérant 10 ms par niveau d'intégration, l'information doit nécessairement se propager de façon proactive, en avant, notamment pour rendre compte des réponses les plus précoces. Si les processus itératifs jouent probablement un rôle dans la perception consciente des sujets, ce rôle pourrait être très limité dans la catégorisation visuelle ultra-rapide. Une troisième et dernière contrainte qui découle de la précédente concerne le nombre de décharges que peuvent émettre les neurones à chaque étape. Toujours avec 10 ms par niveau d'intégration synaptique et un maximum de 100 décharges par seconde et par neurone, à chaque étape, les neurones ont difficilement le temps d'émettre plus d'une décharge. Cette vue séquentielle du traitement visuel est volontairement simpliste et provocatrice mais elle présente l'avantage de poser des questions claires et tranchées. Si, dans ces conditions extrêmes, il est possible de construire des modèles du traitement visuel efficaces, l'adjonction de processus itératifs ne pourra qu'en améliorer les performances.

En général, on considère que l'information visuelle est véhiculée par la fréquence de décharge d'un neurone. Si, à chaque étape, un neurone ne peut émettre qu'une seule décharge, ce type de codage est remis en question. Sur un très court laps de temps, il serait bien sur possible d'estimer une fréquence de décharge sur une population de neurones mais cela nécessiterait un nombre de neurones redondants très important. Une solution alternative, compatible avec la contrainte d'une seule décharge par neurone, propose que l'information soit encodée dans la latence relative d'activation des neurones. Par exemple, pour des neurones ganglionnaires de la rétine encodant une image, la latence de décharge représenterait le contraste au niveau local. Dans une troisième partie, je tente donc d'intégrer ce type de codage temporel et les contraintes énoncées précédemment au sein de modèles simplifiés mais biologiquement plausibles du traitement visuel. Mon but est à la fois de conserver une certaine plausibilité biologique et d'obtenir des performances similaires à celle du système visuel. Je m'attache en particulier à étudier les mécanismes qui pourraient permettre la mise en place d'une sélectivité très rapide des neurones à la fois dans les aires visuelles de bas niveau, où les décharges des neurones sont sélectives à des formes très simples, et dans les aires visuelles de haut niveau, où les décharges des neurones sont sélectives aux objets présentés. Je tente également d'introduire une règle de plasticité synaptique découverte récemment et fondée sur la date relative de décharge des neurones présynaptique et postsynaptique. J'étudie les propriétés émergentes des réseaux de neurones ainsi construits et je m'intéresse en particulier à l'émergence de la sélectivité à l'orientation dans le cortex visuel primaire et l'émergence de la sélectivité aux objets dans les aires corticales visuelles de plus haut niveau. Le modèle le plus évolué que je présente permet notamment de reconnaître des visages dans des photographies. Les performances de ce modèle sont étonnantes du point de vue du traitement d'image et rivalisent aisément avec les approches classiques en intelligence artificielle, notamment en ce qui concerne les capacités de généralisation, de résistance au bruit et de robustesse aux variations de contraste. Les modèles que je présente montrent que, même avec la contrainte drastique d'une propagation des informations visuelles uniquement en avant et pas plus d'une décharge par neurone, il est possible d'atteindre une dimension cognitive de reconnaissance d'objets.

A la fois les expériences et les modèles que j'ai réalisés sont compatibles avec l'hypothèse d'un traitement visuel fondé sur une seule vague de décharges neuronales (il pourrait s'agir de décharge isolée de neurones mais j'indique également pourquoi cette vue est compatible avec une transmission de l'information visuelle par bouffées de décharges). Cependant, je n'affirme pas que cela soit le seul mode de fonctionnement du système visuel. Pour se rapprocher du fonctionnement réel du système visuel, il sera nécessaire d'introduire alors des processus plus complexes, comme des processus itératifs ou des influences descendantes. Cela devrait nous permettre d'atteindre des niveaux cognitifs qui sont restés jusqu'alors réfractaires à la simulation comme l'attention spatiale et pourquoi pas la conscience visuelle. Dans une dernière partie, sur la base des systèmes autopoïétiques, je tente donc de définir les lignes directrices pour construire un modèle computationnel duquel pourrait émerger la conscience.