

SpikeNET : un logiciel pour modéliser de larges réseaux de neurones impulsionnels et asynchrones.

Arnaud DELORME, Rufin VANRULLEN, Jacques GAUTRAIS & Simon THORPE.

Centre de Recherche Cerveau & Cognition,
UMR 5549
Fac. de Médecine Rangueil
133, Rte de Narbonne
31062 TOULOUSE Cedex

Résumé :

Quels sont les mécanismes cérébraux qui sous-tendent la prodigieuse efficacité du système visuel des primates humains et non-humains? Nous avons récemment démontré (Thorpe et al 1996) que le système visuel humain est capable d'analyser une scène visuelle complexe en moins de 150 ms, un niveau de performance qui dépasse très largement les capacités des systèmes artificiels d'analyse d'images.

Ces recherches nous ont permis de dégager plusieurs principes de fonctionnement du système visuel biologique que nous croyons pouvoir compter comme facteurs clés de cette remarquable efficacité. Plus particulièrement, nous sommes convaincus que l'activité impulsionnelle et asynchrone des neurones qui composent le réseau est un élément primordial. Pour tester nos hypothèses, nous avons développé un simulateur de réseaux de neurones impulsionnels que nous avons baptisé "SpikeNET".

Si, au début, le développement de SpikeNET a été surtout motivé par le désir de comprendre le fonctionnement du système visuel humain, nous nous sommes rapidement rendu compte que SpikeNET pouvait être utilisé pour développer des systèmes d'analyse d'images d'un type tout à fait original. SpikeNET a déjà été utilisé pour mettre au point des systèmes de reconnaissance capables de détecter et de localiser des visages humains dans des scènes naturelles.

Introduction.

Des expériences récentes viennent de démontrer très clairement la surprenante efficacité et la rapidité du système visuel (Thorpe et al 1996). Les sujets devaient effectuer une tâche de catégorisation visuelle de type "go/nogo" - répondant à la présentation d'une image uniquement lorsqu'elle contenait un animal. Le niveau de performance obtenu était remarquable (94% correct, avec un temps de réaction médian de 440 ms) malgré (1) la durée de présentation limitée à 20 ms, ce qui empêchait toute possibilité d'exploration oculaire, (2) le fait que chaque image n'était vue qu'une seule fois, ce qui empêchait toute possibilité d'apprentissage, et (3) le fait que les sujets n'avaient aucune information ni sur le type d'animal à chercher (les cibles incluaient des mammifères, des reptiles, des oiseaux, des poissons...), ni sur leur position, leur taille, leur orientation, ou le nombre d'animaux présents. L'utilisation conjointe des potentiels évoqués a de plus permis de démontrer que les réponses cérébrales évoquées lors des essais cibles et des essais distracteurs diffèrent de façon significative dès 150 ms. Le traitement visuel nécessaire pour effectuer cette tâche difficile peut donc être réalisé en moins de 150 ms.

Depuis quelques années, l'universalité de l'hypothèse de codage par fréquence de décharge ("rate coding") est de plus en plus remise en question. Déjà en 1989, nous avons montré

(Thorpe et Imbert, 1989) que la rapidité avec laquelle le système visuel fonctionne est effectivement incompatible avec un traitement basé sur la fréquence de décharge des neurones. L'argument se résume ainsi: on savait que des neurones du cortex inférotemporal pouvaient répondre de façon sélective à certains stimuli complexes comme des visages avec une latence d'environ 100 ms (Perrett et al 1982); on savait également que pour atteindre ces neurones, les signaux générés par les photorécepteurs de la rétine devaient probablement traverser une dizaine de couches de neurones dans la rétine, le thalamus et les diverses aires corticales impliquées dans l'analyse visuelle (V1, V2, V4, PIT, AIT ...), ce qui ne laisse qu'une dizaine de millisecondes par couche pour effectuer le traitement. Etant donné le taux de décharge relativement faible des neurones corticaux (typiquement < 100 Hz), la plupart des neurones n'ont pas eu le temps de générer plus d'un seul potentiel d'action, ce qui pose un problème quasi insoluble pour l'hypothèse de codage par fréquence de décharge car il faut au moins deux impulsions pour pouvoir estimer la fréquence de décharge instantanée. Une estimation de la fréquence de décharge codée sur plusieurs neurones semble également exclue. En effet pour estimer une fréquence de $100\text{Hz} \pm 10\text{Hz}$ en 10ms il faut 281 neurones (Thorpe et Gautrais 1997).

Les neurones intégrateurs à fuite.

C'est en essayant de concilier ces contraintes temporelles avec les connaissances anatomo-physiologiques du système visuel que nous avons été amenés à proposer un nouveau type de codage neuronal basé non pas sur le taux de décharge des neurones, mais plutôt sur leurs dates relatives de décharge (Thorpe et Gautrais 1997).

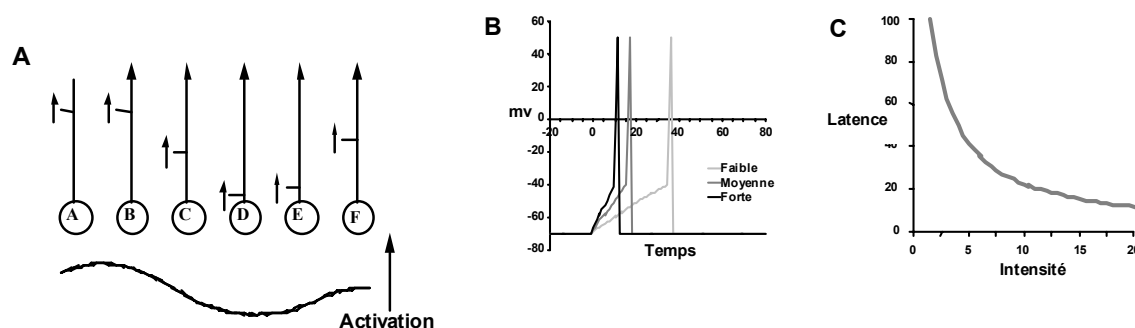


figure 1: A, exemple de conversion d'une activation analogique en latence de décharge des neurones. C'est la latence de décharge des neurones qui code l'information. B, délai d'activation d'un neurone en fonction de l'intensité de la stimulation. C, courbe de correspondance entre la latence de décharge et l'intensité de la stimulation.

Le principe est en fait extrêmement simple, et se base sur l'idée qu'un neurone peut être considéré non pas comme un convertisseur analogique-fréquence comme le veut le dogme neurophysiologique, mais plutôt comme un convertisseur analogique-délai. En effet, on sait que lorsqu'on active un neurone avec un courant constant, le temps nécessaire pour que le potentiel de membrane du neurone atteigne son seuil dépend de l'intensité de l'activation - plus l'activation est forte, plus le neurone va se dépolariser rapidement, et plus courte sera sa latence (figure 1-B,C). La comparaison des dates de décharges de plusieurs neurones peut ainsi coder le stimulus, comme l'illustre la figure précédente (figure 1-A).

Un codage relatif des décharges des neurones offre de plus des caractéristiques intéressantes. Ainsi, le profil d'intensité de la figure 1 peut être représenté par l'ordre

B>A>F>C>E>D. Un des avantages d'un tel type de codage est son invariance par rapport aux changements d'intensité absolue dans une image. En effet, si on double le niveau d'éclairage de l'image ou le niveau de contraste, les latences seront certes raccourcies pour tous les neurones, mais l'ordre dans lequel ils déchargeront restera identique. Ce type d'invariance, caractéristique du système visuel humain, pose de sérieux problèmes aux systèmes classiques d'analyse d'image.

SpikeNET.

SpikeNET est issu de ces travaux. Au départ développé pour étudier le comportement de neurones intégrateurs à fuite décrits précédemment, SpikeNET s'avère être un logiciel capable de rivaliser avec les logiciels d'intelligence artificielle les plus performants.

Chaque élément neuronal est volontairement simplifié, et implique un nombre très réduit de paramètres - un niveau d'activation, un seuil et une sensibilité. Pour éviter tout type de codage fréquentiel, chaque neurone ne peut décharger (émettre un potentiel d'action) qu'une seule fois. Le mécanisme de propagation est également très simple. Le temps est discrétisé, le nombre de pas de temps pouvant correspondre au nombre de neurones afin de minimiser l'incertitude sur l'ordre exact des décharges des neurones. Un neurone, initialement à son potentiel de repos, intègre dans le temps les décharges de ses afférents en modulant leurs poids synaptiques par leur ordre d'arrivée. Cela signifie que le neurone se désensibilise au fur et à mesure qu'il reçoit des décharges de ses différents afférents.

D'un point de vue plus formel, soit $A = \{ a_1, a_2, a_3 \dots a_{m-1}, a_m \}$ l'ensemble des neurones afférents du neurone i , avec $W = \{ w_{1,i}, w_{2,i}, w_{3,i} \dots w_{m-1,i}, w_{m,i} \}$ les poids des m connexions correspondantes; soit f une fonction bornée strictement décroissante telle $\lim_{x \rightarrow +\infty} f(x) = 0 = 0$. Le niveau d'activation du neurone i au temps t est donné par :

$$Activation(i,t) = \sum_{j \in [1,m]} f(ordre_t(a_j)) \cdot w_{j,i}$$

où $ordre_t(a_j)$ représente l'ordre de décharge au temps t du neurone afférent a_j . Par convention, $ordre_t(a_j) = +\infty$ si le neurone a_j n'a pas encore déchargé au temps t , annulant ainsi le terme dans la somme qui correspond à ce neurone.

Le neurone i déchargera au temps t si et seulement si :

$$Activation(i,t) \geq Seuil(i)$$

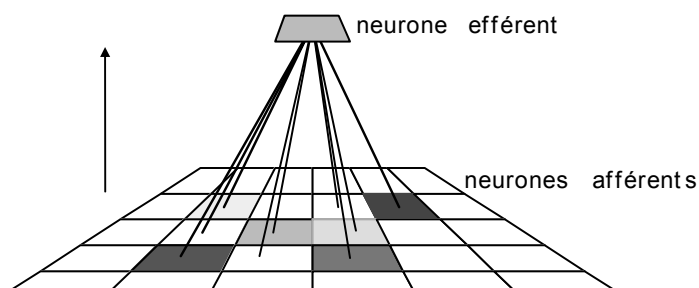


figure 2 : schématisation des projections entre neurones (représentés par des carrés) dans SpikeNet. Les neurones afférents d'un neurone cible sont reliés à celui-ci par des connexions représentées en trait plein. C'est la latence relative d'activation (représentée par les différentes nuances de gris des neurones afférents) qui déterminera, en fonction des poids des connexions et de leur ordre d'arrivée le degré d'activation du neurone cible.

Cette caractéristique fait que SpikeNET ne traite que les neurones qui émettent des potentiels d'action et son efficacité du point de vue computationnel est donc incomparablement supérieure aux réseaux de neurones matriciels classiques.

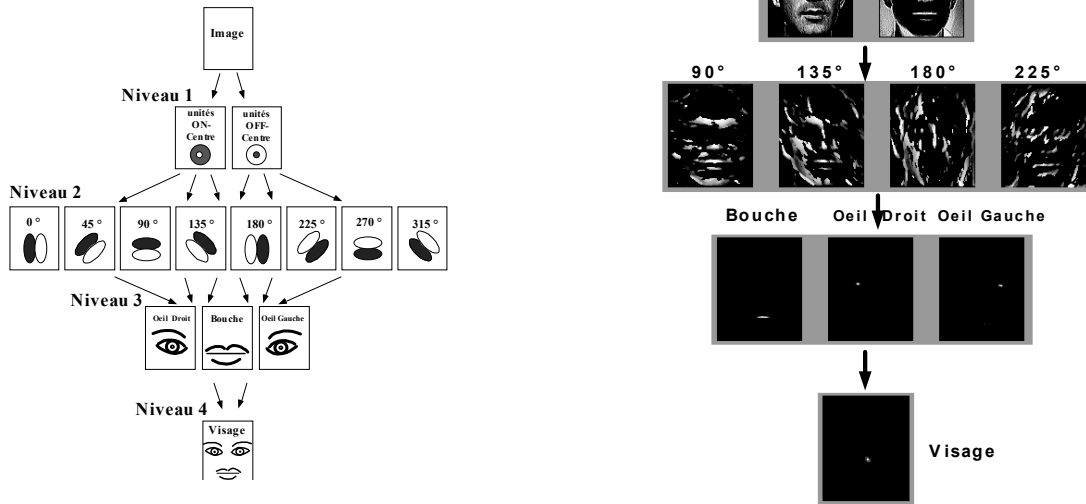
Bien que cela n'ait pas encore été démontré expérimentalement, nous supposons que ce type de comportement intervient au niveau des neurones réels dans le cortex visuel. Des résultats expérimentaux récents semblent indiquer qu'un codage de ce type, c'est-à-dire une dépression du poids synaptique à très court terme, pourrait intervenir au niveau d'une seule synapse (Abbot et al 1997). De plus Markram et Tsodyks (1997) ont montré que lors de la potentiation à long terme d'une synapse d'un neurone cortical, l'efficacité de la synapse se renforce pour le premier potentiel d'action d'un train de potentiels d'action. Il semble donc que pour une synapse dont l'efficacité a été renforcée le premier potentiel d'action soit prépondérant. Enfin Markram et al (1997) ont montré que l'ordre de décharge des neurones intervient pour l'apprentissage au niveau des neurones corticaux : suivant l'ordre de décharge des neurones présynaptique et postsynaptique, l'efficacité des synapses est renforcée ou affaiblie.

Exemple d'application de SpikeNET.

Il s'agit d'un réseau à quatre couches. Le premier niveau correspond à la rétine et contient des cellules dites "ON-centre" et "OFF-centre". Le deuxième niveau contient huit cartes de neurones sélectifs à l'orientation des contours. Le troisième niveau contient des cellules que entraînés à répondre à des caractéristiques spécifiques des visages - l'oeil gauche, l'oeil droit et la bouche. Enfin, le quatrième niveau répond à la présence simultanée de ces trois "caractéristiques" dans des positions appropriées.

La figure suivante illustre comment l'activité se propage à travers un tel réseau. Chaque pixel correspond à un neurone et pour faciliter l'interprétation, l'intensité de ces pixels dépend de l'ordre d'activation des neurones correspondants - plus la date de décharge du neurone est tardive et plus le niveau de gris est sombre. Nous constatons que l'activation dans les cartes correspondant aux yeux et à la bouche permet leur localisation précise, et que, dans la dernière couche, c'est le centre du visage lui-même que l'activité neuronale permet de déterminer. Pour de plus amples explications se référer à VanRullen et al (1998) ou la communication de VanRullen et al pour la détection de visage à plusieurs échelles.

figure 4 : Exemple d'application de SpikeNet. A gauche, l'architecture du réseau de neurones est représentée. A droite le résultat de la propagation d'une image montre que les neurones reconnaissant les visages déchargent quand un visage est présenté au centre de leur champ récepteur.



Conclusion, perspectives.

Nous travaillons actuellement à optimiser les performances de SpikeNET. L'une des caractéristiques essentielles de ce logiciel est que les communications entre les couches de neurones sont extrêmement réduites. SpikeNET est donc tout à fait adapté à une implémentation parallèle, chaque couche de neurone étant traitée par un processeur différent. Etant donné l'absence de tel matériel à un prix abordable, nous collaborons actuellement avec plusieurs entreprises européennes pour développer des cartes multiprocesseurs de type PCI pour Mac ou PC. Le traitement complet pour la recherche de visage dans une image 100*120 prend actuellement 1 seconde sur un Pentium II à 300 Mhz. Cela représente 150 000 neurones et plus de 15 000 000 de connexions. Les cartes multiprocesseurs devraient nous permettre dans les mois qui viennent d'effectuer ce traitement en temps réel sur des images vidéos 640*480. Des applications commerciales de SpikeNET sont également en cours de développement et nous sommes ouverts à tout type de collaboration.

Références

- Abbott L.F., Varela J. A., Kamal Sen, Nelson S. B.** (1997) Synaptic Depression and Gain Control. *Science*, 275, 220-224.
- Markram H., Tsodycks M.** (1997) Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382, 807-810.
- Markram H., Lübke J., Frotscher M.** (1997) Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science*, 275, 213-215.
- Perrett, D. I., Rolls, E. T., & Caan, W.** (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47, 329-342.
- Thorpe, S. J., & Gautrais, J.** (1997). Rapid visual processing using spike asynchrony. In M. Jordan (Ed.), *Neural Information Processing Systems*, 9 (pp. 901-907). Denver.
- Thorpe, S. J., & Imbert, M.** (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreter, F. Fogelman-Souli, & L. Steels (Eds.), *Connectionism in Perspective*. (pp. 63-92). Amsterdam: Elsevier.

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520-522.

VanRullen, R., Delorme, A., Gautrais, J. & Thorpe, S.J. (1998). Face processing using one spike per neurone. *BioSystems*, in press.