

Décodage par IRMf des percepts visuels induits par la lecture de mots

INRIA



La lecture de mots fait intervenir de nombreux processus cognitifs. Afin d'identifier un mot visualisé, le cerveau commence par traiter le percept* visuel, déchiffre les lettres, distingue les bigrammes et active différents mots en se basant sur le contexte ou les attentes, comme la fréquence des mots. Des techniques d'apprentissage automatique vont permettre de décoder la première phase de ce traitement, en utilisant l'*Imagerie par Résonance Magnétique fonctionnelle* (IRMf).

Nous avons établi un décodeur qui prédit le percept visuel formé par des mots de quatre lettres, ce qui nous permet d'identifier des mots qui n'étaient pas présents dans les données d'apprentissage lors de la formation.

**Percept* : forme perçue d'un stimulus externe ou de son absence

Introduction

L'apprentissage automatique peut être utilisé pour prédire, à partir des données obtenues par IRM fonctionnelle (IRMf), une ou plusieurs variables décrivant un stimulus. Cette procédure, souvent désignée par *décodage*, est un élément clé dans la réponse à une question qui se présente naturellement dans les études des processus cognitifs par IRMf : en observant l'activité cérébrale, que peut-on conclure sur l'état mental du sujet accomplissant une tâche particulière ?

Les méthodes de décodage sont largement utilisées pour étudier le système visuel, de l'orientation de textures à la reconnaissance d'objets complexes. Toutefois, afin d'obtenir des informations complètes sur le flot de traitements visuels, une expérience portant sur le décodage doit impérativement aller au-delà de la classification des objets dans des catégories prédéfinies, bien séparées, ou de la prédiction de la forme quel que soit le contenu : elle doit capturer la capacité du cerveau à représenter des stimuli riches [1].

La lecture fait intervenir des tâches cognitives variées, allant du traitement du percept visuel à l'assemblage des lettres et des bigrammes, puis à la formation du mot, et enfin à la résolution de l'ambiguïté entre différents mots sur la base du contexte ou des attentes, comme la fréquence des mots [2]. Une machine capable de faire une déduction similaire en utilisant les données d'IRMf enregistrées, alors qu'un sujet lit le mot, doit tout d'abord capturer l'image dans les aires visuelles précoces.

C'est le problème abordé ici en utilisant des mots de quatre lettres.

Le décodage des mots visuels

Nous nous intéressons à la prédiction des mots présentés à des positions fixes à partir des volumes d'activation observés en IRMf.

Afin de pouvoir résoudre le problème, nous proposons de décrire un mot par un ensemble d'attributs binaires commençant par des descripteurs visuels. En affichant des mots comme sur la Figure 1, chaque lettre est caractérisée par un ensemble de lignes. En éliminant les lettres utilisant des lignes diagonales (par exemple, R), 12 lignes suffisent. Le percept visuel d'un mot de quatre lettres peut alors être décrit par 48 attributs binaires.

Une idée naturelle est alors d'essayer de prédire chaque variable binaire de manière indépendante.



Figure 1.

Un mot de quatre lettres tel qu'il est affiché à l'écran pendant l'expérience. Chaque lettre est assemblée à partir de 12 lignes (6 horizontales et 6 verticales).

Si l'on désigne par f^k la fonction de prédiction de l'attribut $k, 1 \leq k \leq K$ et si l'on suppose en outre que le modèle est linéaire, cela conduit à $f^k(x) = x^T w^k$, qui peut être estimé au moyen d'une fonction logistique.

A partir d'un volume d'IRMf donné, le cadre prédictif fournit la probabilité de chaque attribut composant un mot, $P(y_i^k | X_i)$. En supposant en outre l'indépendance de chaque attribut, on obtient un résultat probabiliste pour un mot entier $y \in \{0,1\}^K : p(y|x_j)$

$$p(y|x_j) = \prod_{k=1}^K p(y_i^k | x_j) \quad (1)$$

Cette procédure fournit un moyen de classer les mots candidats possibles selon les données. Afin d'aller plus loin et de tenir compte de la

structure de corrélation entre les différents attributs, nous avons exploré une approche en deux étapes, dans laquelle un second modèle de régression logistique était appliqué à un nouvel ensemble de variables formé par $\mathbf{XW} \in \mathbb{R}^{N \times K}$.

Résultats

Paradigme : chaque mot était présenté à l'écran pendant 3 s, avec une fréquence de clignotement de 15 Hz. Un intervalle de repos de 5 s était inséré entre chaque présentation de mot. Il était demandé au sujet de fixer une croix colorée située au centre de l'écran. Chaque session comprenait 46 mots, dont 6 verbes. Afin de s'assurer que les sujets lisaient bien les mots, il leur était demandé d'appuyer sur un bouton lorsqu'un verbe était présenté à l'écran. Les répétitions correspondant aux verbes étaient ensuite supprimées de l'analyse. Une carte rétinotopique partiellement utilisant des coins en rotation était obtenue comme dans [3].

Acquisition et traitement IRM : les données d'IRM ont été enregistrées sur 3 sujets en utilisant un système *Magnetom TrioTim 3 Tesla* de Siemens. Les données étaient corrigées en fonction des effets de mouvement. La carte rétinotopique a été calculée pour chaque sujet sur le volume. Pour les besoins de la visualisation, la carte de phase était projetée sur un maillage cortical. La portion aplatie du cortex occipital du sujet 3 est présentée sur la Figure 2. On peut observer les cortex visuels primaires correspondant aux plus grands groupes de vertex actifs le long de la *scissure calcarine* (sillon de la face interne du lobe occipital du cortex).

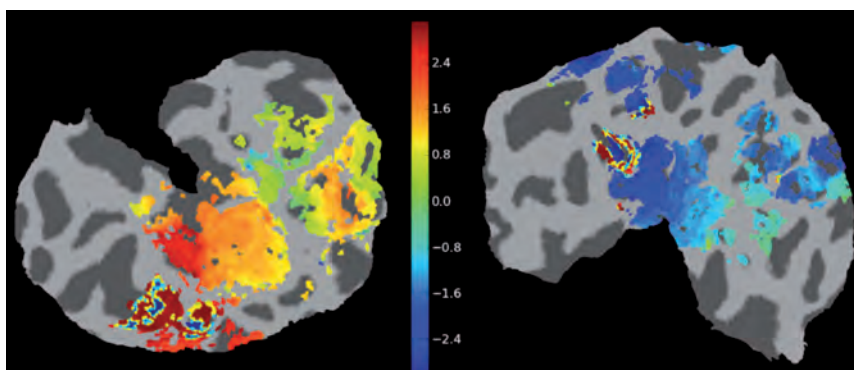


Figure 2.

Cartes de phases rétinotopiques du sujet 3 projetées sur le manteau cortical aplati du lobe occipital (hémisphères gauche et droit). La coupure dans le manteau correspond à la scissure calcarine le long de la paroi interne.

La carte couleur représente l'angle polaire du champ visuel, décalé de $-\pi/2$ et forcé vers l'intervalle $[-\pi, \pi]$.

La classification a été réalisée sur des images cérébrales montrant des estimations de la réponse cérébrale à chaque présentation du stimulus. En affectant à chaque voxel, *pixel en 3D*, l'indice de l'attribut conduisant à l'effet le plus important, il est possible de cartographier chaque voxel selon un attribut, dans le cas présent une ligne parmi les 48 formant un mot. Les résultats sont présentés sur la *Fig.3*.

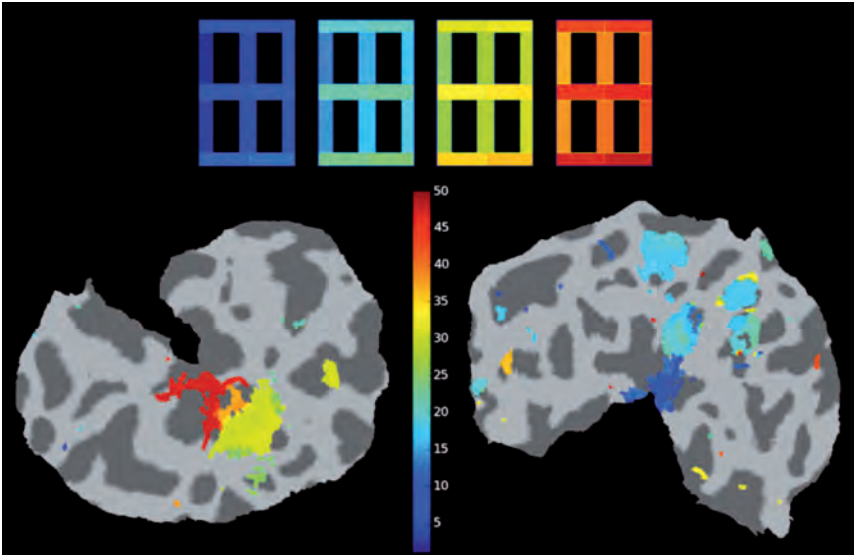


Figure 3.

Les 48 barres visuelles affichées avec un code couleur (du bleu vers le rouge, de la gauche vers la droite) avec les vertex correspondants conduisant aux réponses maximales (seuil du test T $p < 0,001$). Les résultats étaient projetés sur le manteau cortical aplati autour de la scissure calcarine (hémisphères gauche et droit).

On observe des activations essentiellement dans l'aire visuelle V1, avec la symétrie attendue entre le champ visuel gauche (resp. droit) et l'hémisphère droit (resp. gauche). Cette procédure retrouve l'organisation rétinotopique (*Fig.2*), confirmant une bonne fixation par le sujet. On observe également que les lignes situées à la périphérie du champ visuel conduisent à une activation corticale réduite par comparaison avec les régions fovéales. C'est ce que l'on appelle le facteur cortical de grossissement de V1. Cela suggère également qu'un percept d'IRMf plus faible est attendu depuis les lignes périphériques. On peut noter que ces observations sont reproductibles à travers les 3 sujets.

Prédiction des mots : d'après l'équation (1), à partir des probabilités obtenues par les modèles de régression logistique pour chaque ligne, on peut calculer la probabilité d'un mot en fonction des données.

La probabilité peut ensuite être utilisée pour classer les mots dans un corpus. En fonction d'un rang k , cette courbe montre le pourcentage de fois où le vrai mot se situait dans les k premiers mots.

Pourcentage de mots de rang inférieur à k

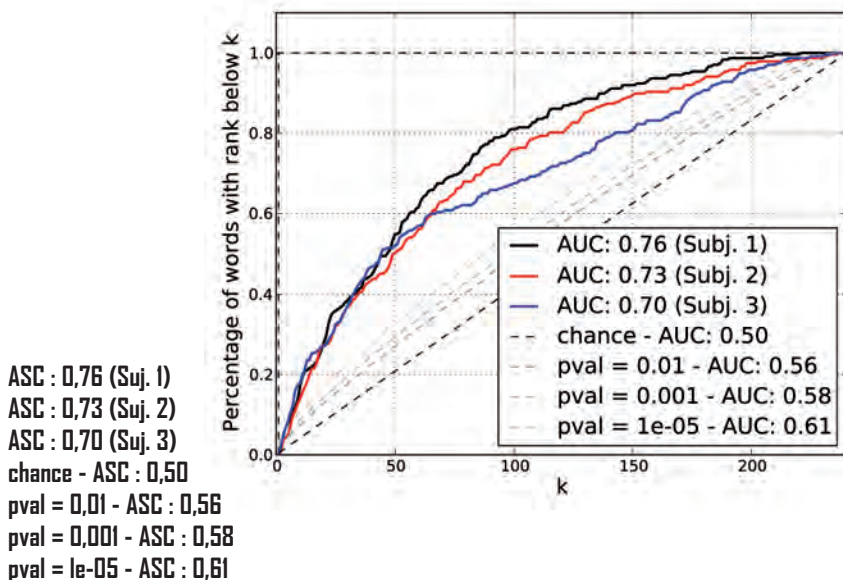


Figure 4. Scores de prédiction des mots en utilisant une mesure de rang pour les 3 sujets, avec un modèle logistique à 2 niveaux. L'ASC obtenue augmente systématiquement par rapport au modèle logistique simple.

Afin de quantifier la performance globale de la méthode, nous calculons un indice d'aire sous la courbe (ASC), qui serait de 0,5 pour des réponses aléatoires. Les résultats obtenus avec le modèle de régression logistique à 2 niveaux, visant à capturer la corrélation entre les attributs visuels, sont indiqués sur la Fig.4.

	suj.1	suj.2	suj.3
score en supprimant 2 mots	95 %	92 %	93 %

Tableau I
Scores de prédiction avec une procédure de validation supprimant deux mots comme dans [1]. Niveau du hasard : 50 %.

Nous avons également supprimé de manière répétée deux mots de l'ensemble de formation et testé si l'algorithme formé sur les mots restants peut prédire quel est le volume de l'IRMf qui correspond à chacun des deux mots restants. Les résultats sont présentés dans le Tableau I.

Conclusion

Dans ce travail, nous avons montré qu'il est possible de prédire le percept visuel induit par un stimulus complexe, comme un mot de quatre lettres, en utilisant des données d'IRMf. Les résultats expérimentaux ont confirmé des observations bien connues en neurosciences concernant l'organisation des aires visuelles primaires.

Travailler avec plusieurs lettres est nécessaire pour mettre au point un algorithme qui extrait les informations pertinentes de données d'IRMf durant le processus de lecture.

Des attributs plus élaborés concernant le mot, comme le contenu sémantique tel qu'il est exploré dans [1], [4], pourraient être utilisés dans de futurs travaux.

Inria

Références pour en savoir plus :

- [1] T. M. Mitchell, S. V. Shinkareva, A. Carlson, K.-M. Chang, V. L. Malave, R. A. Mason, and M. A. Just, *Predicting Human Brain Activity Associated with the Meanings of Nouns*, Science, vol. 320, p. 1191, 2008.
- [2] S. Dehaene, L. Cohen, M. Sigman, and F. Vinckier, *The neural code for written words : a proposal*, Trends in Cognitive Sciences, vol. 9, no. 7, pp. 335 - 341, 2005.
- [3] M. Sereno, A. Dale, J. Reppas, and Kwong, *Borders of multiple visual areas in human revealed by functional magnetic resonance imaging*. Science, p. 889, 1995.
- [4] F. Pereira, G. Detre, and M. Botvinick, *Generating text from functional brain images*, Frontiers in Human Neuroscience, vol. 5, no. 72, 2011.