Université de Montréal

La reconnaissance visuelle à travers le temps

Attentes, échantillonnage et traitement

par

Laurent Caplette

Département de Psychologie

Faculté des Arts et Sciences

Thèse présentée en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en Psychologie, option Sciences Cognitives et Neuropsychologie

Août 2019

Université de Montréal
Département de psychologie, Faculté des Arts et Sciences

Cette thèse intitulée

La reconnaissance visuelle à travers le temps
Attentes, échantillonnage et traitement

Présentée par

Laurent Caplette

A été évaluée par un jury composé des personnes suivantes

Pierre Jolicoeur
Président-rapporteur

Frédéric Gosselin
Directeur de recherche

Karim Jerbi
Codirecteur

Jean-Marc Lina
Membre du jury

Patrick Cavanagh
Examinateur externe

Résumé

La reconnaissance visuelle est un processus temporel : d'abord, l'information visuelle est reçue sur notre rétine de manière continue à travers le temps; ensuite, le traitement de l'information visuelle par notre cerveau prend un certain temps à s'effectuer; finalement, notre perception est toujours fonction autant des expériences acquises dans le passé que de l'input sensoriel présent. Les interactions entre ces aspects temporels de la reconnaissance sont rarement abordées dans la littérature. Dans cette thèse, nous évaluons l'échantillonnage de l'information visuelle à travers le temps pendant une tâche de reconnaissance, comment il se traduit dans le cerveau et comment il est modulé par des attentes spécifiques.

Plusieurs études indiquent que nos attentes modulent notre perception. Comment l'attente d'un objet spécifique influence nos représentations internes demeure cependant largement inconnu. Dans le premier article de cette thèse, nous utilisons une variante de la technique *Bubbles* pour retrouver avec précision le décours temporel de l'utilisation d'information visuelle pendant la reconnaissance d'objets, lorsque les observateurs s'attendent à voir un objet spécifique ou non. Nous observons que les attentes affectent la représentation de différents attributs différemment et qu'elles ont un effet distinct à différents moments pendant la réception d'information visuelle. Dans le deuxième article, nous utilisons une technique similaire en conjonction avec l'électroencéphalographie (EEG) afin de révéler pour la première fois le traitement, à travers le temps, de l'information reçue à un moment spécifique pendant une fixation oculaire. Nous démontrons que l'information visuelle n'est pas traitée de la même manière selon le moment auquel elle est reçue sur la rétine, que ces différences ne sont pas explicables par l'adaptation ou l'amorçage, qu'elles sont d'origine au moins partiellement descendante et qu'elles corrèlent avec le comportement. Finalement, dans le troisième article, nous approfondissons cette investigation en utilisant la magnétoencéphalographie (MEG) et en examinant l'activité dans différentes régions cérébrales. Nous démontrons que l'échantillonnage de l'information visuelle est hautement variable selon le moment d'arrivée de l'information sur la rétine dans de larges parties des lobes occipitaux et pariétaux. De plus, nous démontrons que cet échantillonnage est rythmique, oscillant à

diverses fréquences entre 7 et 30 Hz, et que ces oscillations varient en fréquences selon l'attribut échantillonné.

Mots-clés : vision, perception visuelle, reconnaissance d'objets, reconnaissance de visages, échantillonnage, utilisation d'information, perception temporelle, fréquences spatiales, EEG, MEG.

Abstract

Visual recognition is a temporal process: first, visual information is continuously received through time on our retina; second, the processing of visual information by our brain takes time; third, our perception is function of both the present sensory input and our past experiences. Interactions between these temporal aspects have rarely been discussed in the literature. In this thesis, we assess the sampling of visual information through time during recognition tasks, how it is translated in the brain, and how it is modulated by expectations of specific objects.

Several studies report that expectations modulate perception. However, how the expectation of a specific object modulates our internal representations remains largely unknown. In the first article of this thesis, we use a variant of the *Bubbles* technique to uncover the precise time course of visual information use during object recognition when specific objects are expected or not. We show that expectations modulate the representations of different features differently, and that they have distinct effects at distinct moments throughout the reception of visual information. In the second article, we use a similar method in conjunction with electroencephalography (EEG) to reveal for the first time the processing, through time, of information received at a specific moment during an eye fixation. We show that visual information is not processed in the same way depending on the moment at which it is received on the retina, that these differences cannot be explained by simple adaptation or repetition priming, that they are of at least partly top-down origin, and that they correlate with behavior. Finally, in a third article, we push this investigation further by using magnetoencephalography (MEG) and examining brain activity in different brain regions. We show that the sampling of visual information is highly variable depending on the moment at which information arrives on the retina in large parts of the occipital and parietal lobes. Furthermore, we show that this sampling is rhythmic, oscillating at multiple frequencies between 7 and 30 Hz, and that these oscillations vary according to the sampled feature.

Keywords: vision, visual perception, object recognition, face recognition, sampling, information use, temporal perception, spatial frequencies, EEG, MEG.

Table des matières

Liste des figures

Chapitre 1 : Introduction générale

Chapitre 2 : Article 1

Chapitre 3 : Article 2

Chapitre 4 : Article 3

Chapitre 5 : Discussion générale

Annexe A : Article supplémentaire

Liste des tableaux

Annexe A : Article supplémentaire

Liste des sigles et des abréviations

BOLD : *Blood-Oxygen-Level Dependent*

Bonf. : Bonferroni

CGL : Corps genouillé latéral

COF : Champs oculaires frontaux

Cpd : *Cycle per degree of visual angle*, ou cycle par degré d'angle visuel

Cpf : *Cycle per face*, ou cycle par visage

Cpi : *Cycle per image*, ou cycle par image

Cpo : *Cycle per object*, ou cycle par objet

DNN : *Deep Neural Network*, ou réseau de neurones artificiels profond

EEG : *Electroencephalography*, ou électroencéphalographie

ERP : *Event-Related Potential*, ou potentiel évoqué

FDM : *Frequency-Division Multiplexing*, ou multiplexage fréquentiel

FS : Fréquence spatiale

fMRI : *function Magnetic Resonance Imaging*

FWER : *Family-Wise Error Rate*

FWHM : *Full-Width Half-Maximum*

GLM : *General Linear Model*

IRF : *Impulse Response Function*

IRMf : Imagerie par Résonance Magnétique fonctionnelle

IT : *Inferotemporal Cortex*, ou Cortex inférotemporal

JFI : Jonction Frontale Inférieure

LIP : *Lateral Intraparietal Sulcus*

LOT : *Left Occipito-Temporal*

LTI : *Linear Time-Invariant*

MEG: *Magnetoencephalography*, ou magnétoencéphalographie

MI : *Mutual Information*

MNI : *Montreal Neurological Institute*

PHC : *Parahippocampal Cortex*

RMS : *Root Mean Square*

ROI : *Region of Interest*, ou region d'intérêt

ROT : *Right Occipito-Temporal*

RSC : *Retrosplenial Cortex*

SD : *Standard Deviation*

SF : *Spatial Frequency*

TDM : *Time-Division Multiplexing*, ou multiplexage temporel

Remerciements

Je n'aurais pas pu terminer cette thèse de doctorat sans l'apport et le soutien de nombreuses personnes. D'abord, j'aimerais remercier ma famille et mes amis qui m'ont toujours soutenu dans cette longue et parfois mystérieuse aventure, et qui m'ont permis de passer de bons moments lorsque j'avais besoin de penser à autre chose. J'aimerais particulièrement remercier mes parents qui m'ont maintes fois offert du support et de la nourriture, sans oublier la possibilité d'écrire les derniers chapitres de ma thèse près de la piscine de leur maison! Je ne serais pas rendu où je suis sans le support et la confiance de mes parents et je les remercie énormément.

Au cours de cette longue aventure, j'ai eu la chance de côtoyer plusieurs personnes extraordinaires sans qui le parcours aurait été pénible et difficile. J'aimerais d'abord remercier mon « aîné » de laboratoire Nicolas qui fut un collègue des plus agréables pendant qu'il était encore au laboratoire et qui, encore aujourd'hui, à distance, demeure un ami et un mentor. Merci de m'avoir appris les rudiments des Bulles! Merci également à mes amis et collègues de toujours Jessica et Simon pour les rires, discussions scientifiques et non-scientifiques intéressantes et parfois endiablées, et les conseils pertinents. Vous m'avez fait évoluer comme scientifique et comme personne. J'aimerais aussi remercier Jean-Maxime, plus récemment arrivé au laboratoire, pour son ambition et sa passion : elles m'ont donné encore plus de motivation à devenir le meilleur scientifique que je peux être (et également pour sa folie très divertissante). Merci également à tous les autres que j'ai côtoyés pendant mon long séjour au laboratoire et avec qui j'ai toujours passé de bons moments : Paul, Solène, Alexandre, Xavier, Sandra, Rose-Marie, Valérie. Merci à Mercédès, Sara, et à tous les autres gens du CERNEC avec qui j'ai passé des moments toujours agréables, particulièrement du temps des journées scientifiques à St-Sauveur! Merci également à Ian qui fut dès le début un ami et un mentor des plus généreux. Tes conseils sur l'IRM et la RSA, ta confiance et ton support ont joué un rôle important dans mon parcours. Merci aussi pour l'opportunité de faire une présentation orale à Birmingham. Et merci pour tous les partys auxquels tu m'as invité et pour les gens que tu m'as permis de rencontrer! Merci aussi à Daniel Fiset pour les discussions à VSS et les conseils judicieux. J'aimerais aussi remercier Philippe Schyns pour l'opportunité d'aller

passer quelques mois dans son laboratoire à Glasgow, où j'ai rencontré de merveilleuses personnes, vécu dans un environnement scientifique très stimulant, et appris plein de choses. Merci à Robin pour les bons moments, les discussions intéressantes et l'aide au niveau des analyses. Merci Christoph pour ton amitié qui m'est très importante et pour les discussions scientifiques. Je n'oublierai pas nos discussions dans la salle de pause à Glasgow et les bons moments passés à la fois à Glasgow et à Montréal! Merci également à Caroline Blais et à tous les autres des laboratoires de Ian, Daniel, Caroline et Philippe, pour tous les bons moments à VSS! J'aimerais également remercier Sylvain Baillet de m'avoir donné l'opportunité de faire une présentation orale dans son laboratoire à McGill. Merci aussi à Pierre Jolicoeur pour les cours de doctorat les plus intéressants et pour les conseils toujours pertinents. Merci à Greg West et à Bruno Wicker pour les collaborations. Merci aussi Bruno pour les bons moments à Marseille durant ma maîtrise et les autres moments agréables par la suite. Un grand merci également à Karim Jerbi, mon codirecteur, qui a joué un rôle important dans l'accomplissement de ce doctorat. Merci pour ta confiance, ta générosité, ton caractère un peu fou et tes questions et conseils toujours exceptionnellement pertinents.

Finalement, j'aimerais adresser des remerciements tout particuliers à mon directeur de thèse, Frédéric Gosselin. Merci d'abord de m'avoir fait confiance et d'avoir cru en moi. Grâce à toi, j'ai beaucoup appris et j'ai acquis une confiance plus grande en mes habiletés scientifiques. Merci de m'avoir simultanément accordé une grande liberté et une exceptionnelle disponibilité. J'ai pu faire ce qui m'intéressait le plus et tu as toujours été là pour répondre à mes questions et lire mes (multiples) brouillons d'articles. Peu de professeurs répondent à leurs courriels aussi rapidement et sont au laboratoire aussi fréquemment, et je t'en suis très reconnaissant! Merci pour ton intelligence aiguisée et tes idées originales, et merci pour les réflexions que tu as éveillées en moi. J'aimerais également te remercier pour ton soutien psychologique et tes encouragements dans les moments difficiles, pour toutes les opportunités (présentations, collaborations, conférences, etc.) et pour le soutien financier. Merci également de m'avoir hébergé à San Francisco pendant quelques temps! Par-dessus tout, merci pour les bons moments, les discussions sur le vin et les soirées agréables. Tu as été pour moi plus qu'un simple

superviseur de thèse et j'espère que l'on restera en contact malgré mon départ du laboratoire!

Finalement, cette thèse n'aurait été possible sans le soutien financier du Conseil de Recherche en Sciences Naturelles et en Génie du Canada (CRSNG) et de la Faculté des Études Supérieures et Postdoctorales (FESP) de l'Université de Montréal. Merci également aux Fonds de Recherche du Québec – Nature et Technologies (FRQNT) et à la Vision Sciences Society (VSS).

Chapitre 1

Introduction générale

1.1 Présentation de la thèse

> Rien ne se fait tout d'un coup, et c'est une de mes grandes maximes et des plus vérifiées que la nature ne fait jamais de sauts […].
>
> Gottfried Wilhelm Leibniz, 1765

> Une personne dans une pièce familière […] peu éclairée […] ne pourra identifier aucun objet sinon les plus lumineux […]. Mais tout objet qu'elle reconnaît sera si mêlé à ses souvenirs des meubles se trouvant dans la pièce qu'elle pourra se déplacer dans celle-ci avec sécurité et trouver les objets qu'elle cherche, même s'ils sont peu visibles. Ces images seraient complètement insuffisantes pour reconnaître les objets sans quelque familiarité précédente avec ceux-ci.
>
> Hermann von Helmholtz, 1866

La reconnaissance visuelle est d'une importance fondamentale pour la plupart des animaux. Il s'agit cependant d'un problème computationnel hautement complexe : un même objet peut être vu sous des conditions d'illumination et d'occlusion drastiquement différentes, occasionnant une image rétinienne complètement différente à chaque fois. Malgré cela, nous reconnaissons chaque jour une myriade d'objets animés et inanimés avec très peu d'efforts, sans même y porter attention.

Cette reconnaissance s'opère à travers le temps. D'abord, la stimulation sur notre rétine se déroule dans le temps. Cette stimulation est continue pendant que nous avons les yeux ouverts. De plus, le monde est sans cesse en mouvement et nous sommes constamment en mouvement : l'information entrant dans notre système visuel nous met à jour sur les événements se déroulant dans le monde réel. Ensuite, le traitement de l'information se déroule dans le temps. Le traitement de l'information visuelle dans le cerveau prend un certain temps. La lumière reçue sur la rétine à chaque instant est transformée dans des représentations de plus en plus abstraites dans les différentes régions cérébrales de la hiérarchie visuelle afin de permettre la reconnaissance d'objets. Corollairement, notre conscience est plusieurs millisecondes derrière la réalité, et nous devons souvent recourir à des prédictions inconscientes pour « rattraper la réalité ». Finalement, la reconnaissance s'effectue également à une échelle temporelle beaucoup plus lente : celle de notre vie, pendant laquelle nous acquérons des expériences et créons des

associations entre différents stimuli, associations qui orienteront ensuite le traitement d'autres stimuli avant même leur perception par l'influence des attentes et des a priori.

Dans cette thèse, nous nous attarderons à ces questions d'ordre temporel. Nous nous concentrerons également sur les aspects informationnels de la reconnaissance d'objets : en d'autres mots, nous évaluerons l'utilisation et le traitement d'informations spécifiques, par exemple différentes fréquences spatiales (FS) ou différentes régions d'une image. Nos buts précis et les moyens pour y arriver seront décrits à la fin de cette introduction dans la présentation des articles de la thèse. Avant tout, nous établirons le contexte théorique et méthodologique de la thèse. Nous ferons d'abord un survol sommaire du traitement visuel à travers le temps, puis nous discuterons de la réception et de l'échantillonnage d'information à travers le temps, et du rôle des attentes en perception visuelle; nous poursuivrons avec plusieurs fondations et considérations méthodologiques importantes.

1.2 Le traitement d'objets visuels à travers le temps

Lorsque notre regard s'arrête sur un objet, une partie de la lumière ayant réfléchi sur cet objet entre dans nos yeux et parvient à notre rétine, où elle est transformée en influx nerveux. Ces influx voyagent ensuite à travers le cerveau, du corps genouillé latéral (CGL) aux aires visuelles occipitales et temporales. À chaque étape, l'information véhiculée est transformée et le format de la représentation est modifié. Par exemple, les cellules ganglionnaires de la rétine répondent davantage à des zones lumineuses entourées de zones plus sombres ou à l'inverse (*center-surround cells*). Les cellules de V1 répondent davantage à des barres ou à grilles sinusoïdales de diverses orientations et FS. Les représentations des neurones des aires de plus haut niveau sont plus complexes et élusives, mais les neurones semblent sensibles à des textures de plus en plus complexes. Dans le cortex inférotemporal (IT), là où la reconnaissance semble finalement s'effectuer (DiCarlo, Zoccolan & Rust, 2012; Felleman & Van Essen, 1991; Hong, Yamins, Majaj & DiCarlo, 2016; Lehky & Tanaka, 2016), les neurones semblent sensibles uniquement à un objet donné tout en répondant peu importe son illumination, son orientation, son angle, etc. Il va sans dire que les neurones de IT ont une représentation extrêmement complexe. De plus en

plus, nous parvenons à élucider le format de ces représentations à l'aide de réseaux de neurones artificiels profonds (*Deep Neural Networks* ou *DNNs*; Bashivan, Kar & DiCarlo, 2019; Khaligh-Razavi & Kriegeskorte, 2014; Kriegeskorte, 2015; Yamins & DiCarlo, 2016; Yamins et al., 2014); cependant, la visualisation et la compréhension intuitive de ces représentations demeure ardue. Ces transformations successives peuvent être conceptualisées comme démêlant des surfaces hautement non-linéaires et complexes dans un espace à ultra haute dimensionnalité (Figure 1.1; DiCarlo & Cox, 2007). Plus précisément, nous pouvons conceptualiser la représentation d'un objet par une région cérébrale comme une surface non-linéaire (*manifold*) dans un espace cartésien avec un nombre de dimensions correspondant au nombre de neurones de cette région. Ces surfaces sont hautement non-linéaires dans les régions de bas niveau, puisque deux objets peuvent se présenter sous de multiples formes et occasionner des patterns d'activité très différents; ainsi, séparer les surfaces représentant différents objets peut s'avérer particulièrement difficile. Dans les régions de plus haut niveau, ces surfaces deviendraient plus lisses, ce qui permettrait de séparer les représentations de différents objets (i.e. de décoder l'identité d'un objet) plus facilement.

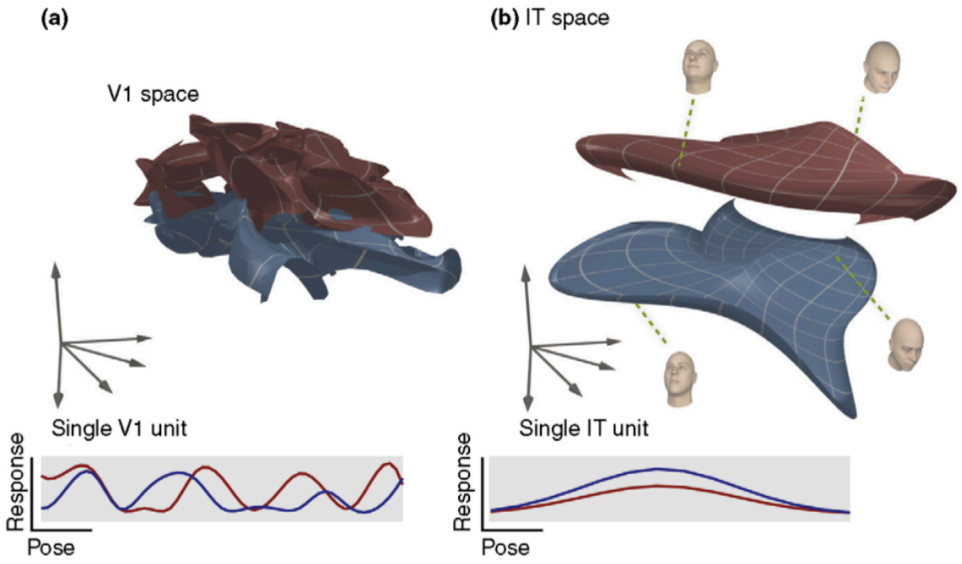


Figure 1.1. Illustration de surfaces représentant deux objets (dans ce cas-ci deux visages d'identités distinctes) dans des espaces neuraux abstraits. Dans V1, ces surfaces sont hautement non-linéaires et difficilement séparables; dans IT, ces surfaces sont beaucoup plus lisses et peuvent être séparées à l'aide d'une simple frontière décisionnelle linéaire. Figure adaptée de DiCarlo & Cox (2007).

Ces transformations s'effectuent évidemment à travers le temps, les différentes régions cérébrales répondant plus ou moins une à la suite de l'autre. Plusieurs études ont démontré que, chez le primate, les neurones de V1 déclenchent après environ 50 ms, ceux de V2 déclenchent après environ 60 ms, ceux du cortex inférotemporal (IT) après environ 80 ms et ceux du cortex latéral intrapariétal (LIP) après environ 100 ms (Figure 1.2; Bullier, 2001; DiCarlo et al., 2012). Plus récemment, il a été démontré que l'activité dans V1 permet de décoder maximalement l'identité des objets autour de 101 ms alors que celle de IT le permet autour de 132 ms. De plus, le format de la représentation dans V1 semble être

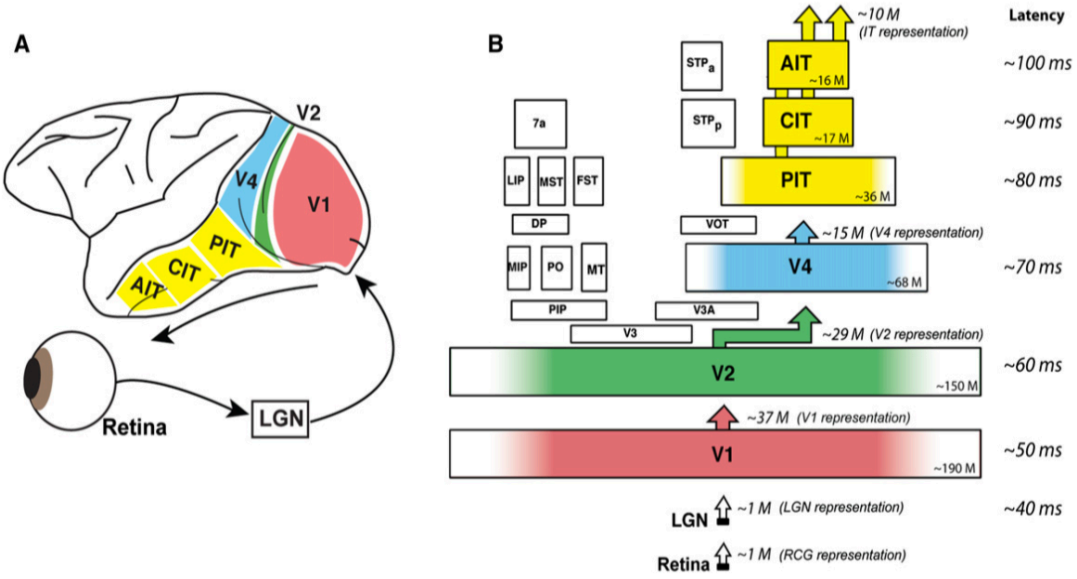


Figure 1.2. Illustration des principales régions cérébrales en jeu lors de la reconnaissance d'objets chez le macaque. Figure tirée de DiCarlo et al., 2012.

relativement semblable pendant les 600 premières millisecondes, alors que la représentation dans IT est en constante évolution pendant ce temps (Cichy, Pantazis & Oliva, 2014).

En électroencéphalographie (EEG) et en magnétoencéphalographie (MEG), qui échantillonnent l'activité neuronale de manière indirecte en évaluant l'activité électromagnétique sur le scalp, ce traitement temporel de l'information visuelle est visible (lorsqu'on moyenne à travers plusieurs essais) par une succession de composantes, c'est-à-dire par une série d'inflexions positives ou négatives dans le voltage ou le champ magnétique (Figure 1.3). Ces composantes ont également différentes répartitions sur le

scalp. Les composantes les plus fréquemment observées lors du traitement d'un stimulus visuel sont (en EEG) : la C1 (maximum autour de 90 ms), associée au traitement dans les aires visuelles primaires et changeant de polarité selon la localisation du stimulus dans le champ visuel; la P1 (100-130 ms), associée au traitement dans le cortex occipital extrastrié et sensible à l'attention du sujet; la N1 (150-200 ms), également sensible à l'attention du sujet et visible sur les électrodes frontales avant les électrodes plus postérieures; la N170 (130-200 ms), évoquée par l'apparition de visages et dont la topographie est différente de la N1; la P2 (150-275 ms), associée à un traitement visuel de niveau intermédiaire; et la P3 (250-500 ms), semblant refléter des processus associés à l'évaluation et à la catégorisation des stimuli.

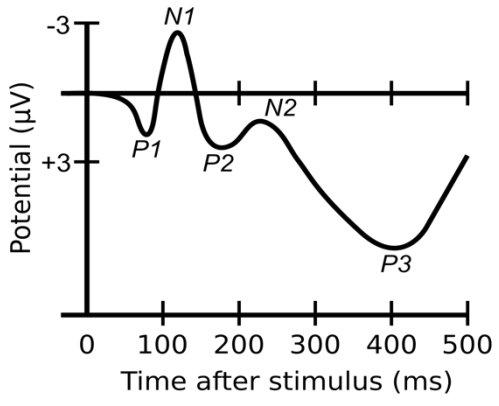


Figure 1.3. Illustration des principales composantes EEG associées au traitement visuel.

1.3 La réception et l'échantillonnage d'information visuelle à travers le temps

Afin de reconnaître un objet, les photorécepteurs de notre rétine doivent recevoir une certaine quantité de photons qui ont réfléchi sur cet objet : l'objet doit ainsi être fixé pendant plus qu'un instant. Cette nécessité pourrait résulter du fait que les signaux neuronaux sont bruités, obligeant ainsi une certaine redondance dans l'information reçue. Cela pourrait également être causé parfois par un changement dans l'information parvenant à la rétine à travers le temps, soit parce que l'objet est dynamique (e.g., un visage humain démontrant une expression faciale dynamique) ou simplement parce que l'observateur bouge les yeux. Une autre possibilité est la présence d'un goulot d'étranglement, attentionnel ou autre, dans les aires visuelles de plus haut niveau (voir section 1.3.1.2) : si seulement une quantité limitée d'information peut être traitée à la fois, déplacer l'attention sur différents attributs à différents moments est un moyen de compenser cette limitation, mais le traitement de l'objet nécessiterait alors une fixation prolongée.

Le fait qu'un objet doit être fixé pendant plus qu'un instant, et qu'il l'est normalement pendant 200 à 300 ms (la durée typique d'une fixation oculaire), implique que l'information visuelle (i.e. la lumière ayant réfléchi sur cet objet) reçue sur la rétine à différents moments est simultanément traitée dans le cerveau, possiblement à différents niveaux (i.e. l'information reçue il y a longtemps est probablement traitée à un stade plus avancé pendant que l'information reçue il y a quelques millisecondes est traitée à un stade moins avancé; Figure 1.4). Ainsi, la réponse cérébrale à un stimulus telle que typiquement évaluée est en fait une somme de réponses cérébrales à des informations reçues à différents moments et les stimuli, mêmes statiques, ne peuvent pas être considérés comme simplement spatiaux car ils sont intrinsèquement spatiotemporels.

1.3.1 Traitement différent de l'information reçue à différents moments

L'information reçue à différents moments sur la rétine risque d'être traitée différemment, même pendant une fixation oculaire et avec une information sensorielle constante (Figure 1.4b). Il y a plusieurs raisons possibles à une telle variation dans le traitement.

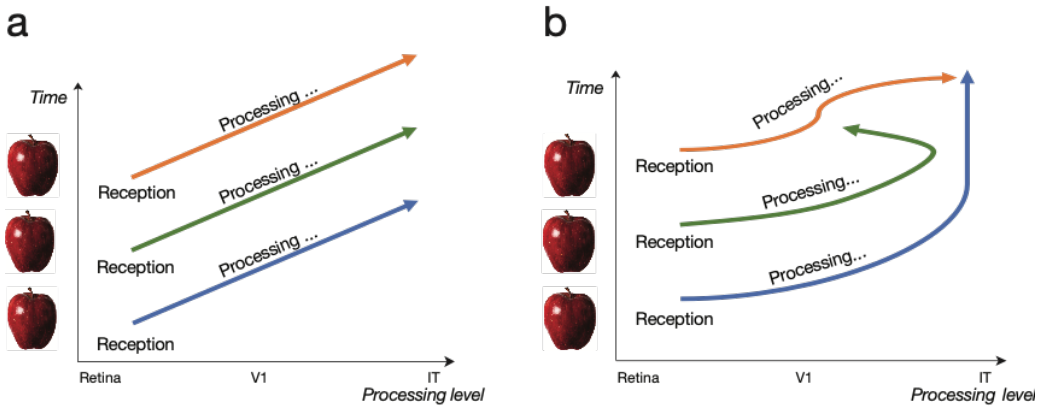


Figure 1.4. À n'importe quel point dans le temps (n'importe quelle ligne horizontale imaginaire dans les diagrammes ci-dessus), l'information reçue à différents moments durant la fixation est simultanément traitée dans le cerveau, possiblement à différents stades de traitement. A) Le traitement de l'information reçue à différents moments est identique. B) Le traitement de l'information reçue à différents moments est différent. Figure reprise de l'article #2 de cette thèse.

1.3.1.1 Mécanismes de bas niveau

D'abord, des mécanismes de bas niveau tels que l'adaptation et l'amorçage sont susceptibles d'altérer le traitement de l'information reçue au début d'une fixation ou à sa fin. En effet, la réponse d'un neurone peut être réduite ou augmentée suite à la répétition d'un stimulus ou à sa présentation prolongée.

Également, la présence d'oscillations cérébrales risque de faire en sorte que l'information reçue à certaines phases de ces oscillations soit privilégiée par rapport à l'information reçue à d'autres phases. La phase pré-stimulus des oscillations cérébrales dans plusieurs bandes de fréquences a été associée à des fluctuations dans le seuil perceptif (Busch & VanRullen, 2010; Helfrich et al., 2018). Similairement, des oscillations ont été observées dans plusieurs variables comportementales telles que l'utilisation d'information, l'attention et le seuil de détection (Blais, Arguin & Gosselin, 2013; Latour, 1967; VanRullen, Carlson & Cavanagh, 2007; voir VanRullen, 2016, pour une revue de la littérature).

1.3.1.2 Sélection descendante d'information à travers le temps

Une autre possible cause de variation est le déplacement couvert (*covert*) de l'attention à travers le temps. Les humains ont une capacité cognitive intrinsèquement limitée : il y a un goulot d'étranglement dans les régions de plus haut niveau (Cowan, 2000; Desimone & Duncan, 1995). Une première stratégie pour remédier à cette limitation pourrait être de maintenir toute l'information dans les régions de bas niveau (dans un tampon) et de transmettre l'information en paquets différés aux aires supérieures. Cependant, cette stratégie repose sur la supposition que les aires inférieures ont une capacité suffisante pour le maintien de toute cette information. De plus, cette stratégie est potentiellement sous-optimale dans l'éventualité où l'information visuelle changerait durant la fixation. En effet, à moins d'impliquer des mécanismes supplémentaires, les aires supérieures n'auraient pas accès à l'information la plus à jour dès que possible. Ainsi, une stratégie plus sensée est de traiter l'information reçue sur la rétine à différents moments de manière à y traiter différents attributs – en d'autres mots, de déplacer l'attention (soit spatialement ou dans un espace plus abstrait) à travers le temps.

Cette sélection attentionnelle peut survenir dans presque toutes les régions cérébrales (Desimone & Duncan, 1995) et aussi tôt que le CGL (McAlonan, Cavanaugh & Wurtz, 2008). La source du signal descendant dirigeant l'attention aux cibles appropriées semble être les champs oculaires frontaux (COF) dans le cas de l'attention spatiale (Gregoriou, Gotts, Zhou & Desimone, 2009) et la jonction frontale inférieure (JFI) dans le cas de l'attention non-spatiale (Baldauf & Desimone, 2014). Ces régions frontales communiqueraient avec le cortex visuel via une synchronisation des oscillations gamma (Baldauf & Desimone, 2014; Gregoriou et al., 2009).

L'attention semble aussi osciller à travers le temps. Ces fluctuations semblent survenir à une fréquence d'environ 7-8 Hz (Busch & VanRullen; 2010; Holcombe & Chen, 2013; VanRullen et al., 2007; VanRullen, 2016; Verstraten, Cavanagh & Labianca, 2000). Même s'il y a un seul objet, cette oscillation est présente; s'il y a deux objets dans le focus attentionnel, l'attention sur chaque objet oscille à une fréquence de 3-4 Hz, suggérant que la moitié des cycles d'une oscillation attentionnelle à 7-8 Hz est dédiée à chaque objet en alternance (Fiebelkorn, Saalmann & Kastner, 2013; Fiebelkorn, Pinsk & Kastner, 2018; Landau & Fries, 2012).

Plutôt que d'utiliser un multiplexage temporel (i.e. dédier différents moments à différents attributs en alternance), un multiplexage fréquentiel pourrait être utilisé : différents attributs peuvent ainsi être codés à différentes fréquences, permettant leur codage parallèle. Ce phénomène a été observé au niveau du traitement cérébral (Smith, Gosselin & Schyns, 2006; Romei, Driver, Schyns & Thut, 2011), à la fois pour de l'information spatiale et pour de l'information non-spatiale. Cependant, les preuves empiriques de multiplexage fréquentiel dans l'échantillonnage sont très limitées au niveau du comportement (voir Dupuis-Roy, 2014) et inexistantes au niveau du cerveau.

Une utilisation[1] séquentielle non-oscillatoire d'attributs pourrait également avoir lieu parce qu'il s'agit d'une stratégie plus efficace que d'utiliser n'importe quel attribut à

[1] Nous employons ici le terme « utilisation » plutôt que traitement (même si l'un ne va évidemment pas sans l'autre) afin de bien mettre au clair que nous faisons allusion à la dimension temporelle de la réception d'information sur la rétine. Ainsi, quand nous indiquons « L'attribut X est utilisé avant l'attribut Y », nous voulons signifier que l'attribut X reçu tôt est traité alors que l'attribut Y est traité seulement lorsqu'il est reçu plus tard. Évidemment, cette manière de présenter les choses repose sur une conception binaire de l'utilisation : il est possible qu'un attribut reçu à un moment X élicite une activité neuronale d'une certaine

n'importe quel moment (Cavanagh, Labianca & Thornton, 2001; Cavanagh, 2004; Jolicoeur, Ullman & Mackay, 1986, 1991; Ullman, 1984). Par exemple, traiter l'information en basses fréquences spatiales (FS) avant celle en plus hautes FS serait computationnellement plus efficace lors de la reconnaissance d'objets et de scènes (Marr, 1982; Watt, 1987). Une utilisation des basses FS débutant plus tôt que l'utilisation des hautes FS a effectivement été démontrée en reconnaissance d'objets (Caplette, Wicker & Gosselin, 2016; Caplette, Wicker, Gosselin & West, 2017a; Hughes, Nozawa & Kitterle, 1996; Parker, Lishman & Hughes, 1996) et de scènes (Schyns & Oliva, 1994). D'autres biais pourraient résulter en des stratégies séquentielles stables chez plusieurs individus : par exemple, une tendance à échantillonner[2] les attributs les plus informatifs d'abord ou une tentative de compenser des limitations anatomiques (échantillonner plus tôt un attribut qui prend davantage de temps à être traité; e.g., Dupuis-Roy et al., 2019). Il est à noter que ces stratégies, ou routines visuelles (Ballard, 2015; Cavanagh, 2004; Cavanagh et al., 2001; Ullman, 1984), risquent fort de dépendre des stimuli attendus et de la tâche à effectuer (Schyns & Oliva, 1999).

Pour l'instant, nous avons surtout fait allusion à des routines visuelles balistiques : ces routines d'échantillonnage sont prédéterminées avant la fixation de l'objet et demeurent ensuite immuables. Cependant, une routine visuelle pourrait également être adaptative (Caplette et al., 2017b). Dans un tel scénario, l'information reçue et traitée tôt pourrait influencer l'échantillonnage et le traitement de l'information reçue plus tard, et ce, malgré un input visuel qui reste constant[3]. Une telle routine nécessiterait probablement un traitement descendant, l'information reçue tôt devant être traitée jusqu'à un certain point afin de pouvoir moduler le traitement de l'information reçue plus tard. Comme un tel traitement prend un certain temps, une telle modulation ne pourrait pas survenir avant potentiellement quelques centaines de millisecondes. C'est ce qui a mené certains auteurs

amplitude alors que le même attribut reçu plus tard élicite une activité neuronale d'une moindre amplitude – il est également possible que le traitement soit qualitativement différent d'un moment de réception à l'autre. C'est pourquoi ces termes seront surtout employés lorsque nous discutons du comportement (plutôt que du traitement cérébral), pour garder le texte simple et concis.

[2] Le terme échantillonnage sera employé de manière similaire au terme utilisation (*utilisation* a cependant un sens plus restreint : voir section 1.5.1).

[3] Notons qu'un traitement peut être adaptatif sans que l'échantillonnage le soit. En effet, on pourrait imaginer que l'information reçue au même moment mais traitée plus rapidement soit utilisée afin d'influencer le traitement subséquent de l'information qui est traitée plus lentement.

à suggérer que seule l'attention ouverte (*overt*; i.e. le déplacement des yeux) est sujette à un traitement descendant et qu'une routine visuelle est balistique à l'intérieur d'une fixation (Ballard, 2015). Également, il est possible qu'une routine soit balistique sur une dimension et adaptative sur d'autres : par exemple, les FS pourraient être extraites des basses aux plus hautes en tout temps, mais l'endroit où les hautes FS sont échantillonnées dans le champ visuel pourrait être déterminé par le contenu de l'information en basses FS. Même si l'adaptabilité est certainement présente sous une certaine forme et à une certaine échelle temporelle chez l'humain, il est loin d'être certain qu'elle soit présente durant la reconnaissance même d'un objet. En effet, même s'il semble évident qu'utiliser l'information extraite tôt est un avantage, cela pourrait requérir davantage de temps et de ressources computationnelles et atteindre le même taux de reconnaissance pour une majorité de situations.

1.3.1.3 Accumulation et intégration de l'information visuelle

Puisque l'information provenant de différents moments durant la fixation doit être intégrée en vue de prendre une décision sur l'identité d'un objet, l'information reçue à différents moments doit être accumulée avant la reconnaissance. Ainsi, l'information reçue tôt doit être maintenue plus longtemps que l'information reçue plus tard : leur traitement, à travers le temps, devrait donc être différent. Les corrélats cérébraux de l'accumulation d'information dans le cerveau ne sont toujours pas entièrement déterminés. Les psychologues spéculent depuis longtemps sur la nature discrète ou continue de la transmission d'information dans le cerveau, c'est-à-dire si le traitement à un stade donné doit être complété dans son entièreté avant que l'information soit transférée au stade suivant (ce qui nécessiterait l'accumulation d'information) ou si les résultats partiels sont graduellement transmis (Miller, 1988). À tout le moins, le dernier niveau de traitement doit nécessairement effectuer une certaine forme d'accumulation, si l'information reçue durant toute la fixation est prise en compte. La composante de potentiels évoqués visuels P3 pourrait correspondre à un tel processus d'accumulation (Twomey, Murphy, Kelly & O'Connell, 2015). Les COF et le LIP ont également été proposées comme régions responsables de l'accumulation d'information et plusieurs études supportent ces

hypothèses (pour une revue de la littérature, voir Hanks & Summerfield, 2017; Huk, Katz & Yates, 2017).

1.3.2 Traitement vs échantillonnage

Des différences temporelles dans le traitement de différents attributs d'un stimulus (e.g., Schoenfeld, Hopf, Merkel, Heinze & Hillyard, 2014; Smith, Fries, Gosselin, Goebel & Schyns, 2009) peuvent être causées par deux phénomènes distincts : soit les attributs ont été traités à différentes vitesses, soit ils ont été échantillonnés à différents moments (i.e. l'attention a été portée sur ceux-ci à différents moments). Sans information supplémentaire, la situation est intrinsèquement ambiguë.

Deux attributs pourraient être traités à différents moments dans les aires visuelles de haut niveau si des chemins neuronaux différents sont employés pour les traiter. Différents chemins peuvent être constitués d'un nombre de relais (et de neurones) différents (et de manière reliée, pourraient mettre en jeu des opérations prenant des temps différents), ou pourraient simplement être constitués d'axones plus petits ou moins myélinisés transmettant l'information plus lentement. Alternativement, ces attributs pourraient être traités à différents moments simplement parce qu'ils ont été échantillonnés à des moments différents. En d'autres mots, si un attribut n'atteint la région cérébrale qu'après un autre, c'est peut-être parce que l'information reliée à cet attribut provenant du début de la fixation n'a tout simplement *pas* atteint la région cérébrale, potentiellement parce que l'information a cessé d'être représentée avant d'atteindre la région (i.e. l'attention couverte n'a pas été dirigée sur cet attribut au début de la fixation).

Ainsi, le moment de traitement d'un attribut dans une aire corticale donnée est fonction à la fois du moment d'échantillonnage (i.e. le traitement est effectué sur l'information reçue sur la rétine à quel moment?) et de la vitesse de traitement (i.e. le traitement de la rétine à cette région cérébrale a pris combien de temps?). Ces concepts sont rarement discutés et souvent confondus dans la littérature; cela peut mener à des interprétations en termes de traitement lorsqu'il s'agit d'échantillonnage ou vice-versa (voir VanRullen, 2011). Par exemple, on sait que les cellules parvo, qui traitent

l'information chromatique, conduisent l'information plus lentement que les cellules magno, qui traitent l'information achromatique (différentes *vitesses de traitement*); ainsi, les cellules répondant à la couleur dans V2 sont activées 10-20 ms plus tard que les cellules ne répondant pas à la couleur lorsque les stimuli sont de simples points lumineux (Nowak, Munk, Girard & Bullier, 1995; différents *moments de traitement*). Cependant, on observe également que, dans une tâche de catégorisation du genre de visages, les observateurs utilisent l'information chromatique avant la majorité de l'information achromatique (Dupuis-Roy, Faghel-Soubeyrand & Gosselin, 2019; différents *moments d'échantillonnage*). Même si ces résultats peuvent paraître contradictoires, nous avons vu comme quoi les moments de traitement et d'échantillonnage sont deux concepts différents qui ne sont pas nécessairement corrélés. Dans ce cas, il se pourrait que le cerveau échantillonne l'information chromatique avant l'information achromatique durant la catégorisation du genre afin de s'assurer qu'elle va atteindre les régions visuelles de haut niveau approximativement au même moment que l'information chromatique (qui est traitée plus lentement).

De plus, on sait que la voie magnocellulaire achemine les basses FS de la rétine au CGL plus rapidement que la voie parvocellulaire qui traite les hautes FS (Bullier & Nowak, 1995; Nowak et al., 1995) et qu'un traitement plus rapide des basses FS a également été rapporté dans le CGL (Allen & Freeman, 2006) et dans V1 (Mazer, Vinje, McDermott, Schiller, & Gallant, 2002; Purushothaman, Chen, Yampolsky, & Casagrande, 2014). D'autre part, des observateurs neurotypiques échantillonnent typiquement les basses FS avant les plus hautes dans une tâche de reconnaissance d'objets ou de scènes (Figure 1.5; Caplette et al., 2016, 2017a; Schyns &

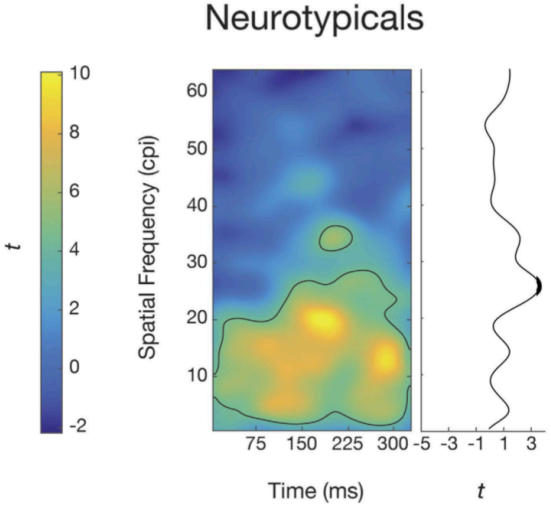


Figure 1.5. Illustration de l'utilisation des FS à travers le temps chez des sujets neurotypiques. Les contours noirs indiquent l'utilisation significative d'information. Le graphe 1D sur le côté indique à quel point l'utilisation des différentes FS augmente à travers le temps : la section en gras indique une augmentation significative. Figure adaptée de Caplette et al. (2016).

Oliva, 1994). Même si ces résultats peuvent sembler reliés, ils ne le sont pas : les premiers réfèrent à la vitesse de traitement (qui est probablement essentiellement immuable) tandis que les derniers réfèrent au moment d'échantillonnage (qui peut être largement influencé par la tâche et le contexte). D'ailleurs, Oliva & Schyns (1997) ont démontré que des participants peuvent être entraînés à échantillonner les hautes FS avant les plus basses dans une tâche d'identification de scènes (voir aussi Schyns & Oliva, 1999). On peut également observer que les FS et les temps en jeu sont drastiquement différents entre ces deux groupes d'études : latences de 28 à 70 ms et FS de 0 à 21 cycles par degré pour le traitement des FS par les cellules magno et parvo (Mazer et al., 2002), et moments de 0 à 333 ms et FS de 0 à 6 cycles par degré pour les FS utilisées dans la reconnaissance d'objets (notons que l'échelle temporelle en jeu ici est arbitraire et dépend des stimuli utilisés, voir section 5.5.1; Caplette et al., 2016).

En utilisant des stimuli statiques, il faudrait connaître le chemin complet emprunté par l'information visuelle de la rétine à la région d'intérêt, avec une haute résolution spatiale et temporelle, pour départager les influences de la vitesse de traitement et du moment d'échantillonnage. En effet, il faudrait connaître le moment précis de l'arrivée de l'information sur la rétine et ensuite suivre cette information d'une région à l'autre jusque dans la région d'intérêt, et faire de même pour l'information arrivée à chaque moment sur la rétine; autrement, nous ne pourrions être certain si l'arrivée plus tardive d'un attribut par rapport à un autre est due à un échantillonnage plus tardif ou à une transmission plus lente. Cependant, étant donné les limitations méthodologiques actuelles, une telle chose est très difficile, voire impossible. D'abord, nous n'avons pas accès à la rétine, point d'entrée de l'information dans le système visuel, à l'aide des techniques de neuroimagerie permettant d'imager l'ensemble du cerveau : cela est problématique car les différentes informations ne sont pas toutes transmises à la même vitesse de la rétine aux premiers relais sous-corticaux (e.g., basses FS vs hautes FS; Nowak et al., 1995). Ainsi, nous ne pourrions pas supposer que deux attributs traités au même moment dans les régions suivant la rétine ont bien été échantillonnés au même moment. Similairement, nous n'avons pas non plus accès aux premiers relais sensoriels sous-corticaux tels que le CGL; il a cependant été démontré qu'une première sélection d'information peut s'y effectuer (O'Connor, Fukui, Pinsk & Kastner, 2002; McAlonan et al., 2008). Finalement, même au niveau du cortex, la

résolution spatiale des techniques d'imagerie de l'ensemble du cerveau est possiblement trop faible (quoique cela reste à déterminer) pour avoir accès à tous les nœuds fonctionnels où l'information peut être sélectionnée et ainsi, il serait difficile de suivre une information reçue à un moment précis sans la confondre dans une région plus tardive avec une information reçue à un autre moment. Nous proposons une solution à ces problèmes en utilisant des stimuli dynamiques et une technique de corrélation inverse (voir section 1.5).

1.4 L'effet d'attentes préalables sur la reconnaissance visuelle

Les êtres humains construisent sans cesse des modèles d'eux-mêmes et de leur environnement afin de prédire le monde avec le plus d'exactitude possible (de Lange et al., 2018). Certains postulent même que le cerveau serait constamment en train de tenter de prédire l'environnement de manière descendante et que seule la partie erronée de cette prédiction serait transmise de manière ascendante (plutôt que toute l'information sensorielle; Clark, 2013; Friston, 2010). L'attente d'un objet spécifique altère notre perception de cet objet, s'il apparaît par la suite. Nous détaillons ces effets des attentes dans la prochaine section.

1.4.1 L'effet des attentes sur la perception

Lorsque l'information sensorielle est faible, bruitée ou ambiguë, l'attente d'un stimulus peut biaiser notre perception. Par exemple, la direction dans laquelle on perçoit bouger un nuage de points sera influencée par notre attente si le contraste ou la cohérence du mouvement sont faibles (Chalk, Seitz & Seriès, 2010; Kok, Brouwer, van Gerven & de Lange, 2013). Également, si les points bougent dans des directions complètement aléatoires et qu'il n'y a donc aucune cohérence de mouvement, des sujets s'attendant à percevoir un mouvement précis peuvent percevoir le nuage de points comme bougeant dans cette direction (Sterzer, Frith & Petrovic, 2008). De plus, si on présente uniquement des images de bruit à un sujet et qu'on lui indique qu'une image est parfois présente derrière ce bruit,

il percevra cette image : il est possible de retrouver sa représentation interne (Gosselin & Schyns, 2003; voir section 1.5.2.1).

Lorsque les stimuli ne sont pas du tout ambigus ou lorsque nos attentes sont complètement éloignées de la réalité, les effets des attentes risquent d'être moins prononcés et de se réduire à des temps de réponse plus faibles (Pinto, van Gaal, de Lange, Lamme & Seth, 2015; Stein & Peelen, 2015) et à une exactitude plus élevée (Wyart, Nobre & Summerfield, 2012a) lorsque les attentes sont valides. Ces effets ont été documentés également avec des objets complexes réalistes. Lorsque l'on s'attend à percevoir un objet spécifique et que celui-ci apparaît, à la fois sa détection (Stein & Peelen, 2015) et sa catégorisation (Esterman & Yantis, 2010; Puri & Wojciulik, 2008) s'en trouvent améliorées.

De manière générale, les observateurs vont se fier davantage à leurs attentes, ou *a priori*, lorsque les stimuli sont ambigus et ils vont se fier davantage à l'information sensorielle lorsque les attentes ne sont pas fiables. Ainsi, les attentes et l'information sensorielle semblent pondérées par leur incertitude; une telle pondération est en accord avec les postulats de modèles de perception Bayésienne et de codage prédictif (de Lange, Heilbron & Kok, 2018; Kersten, Mammassian & Yuille, 2004; voir section 1.4.3).

1.4.2 Origines des attentes

Les attentes peuvent se présenter sous plusieurs formes et avoir plusieurs origines. Certaines « attentes », ou a priori, résultent des régularités dans les statistiques de bas niveau de notre environnement. Par exemple, nous sommes plus sensibles aux orientations cardinales, qui sont plus présentes dans notre environnement. De plus, nous tendons à percevoir la lumière comme provenant d'en haut, ce qui fait en sorte que l'on perçoit certaines formes comme concaves alors qu'elles pourraient tout aussi bien être convexes si la lumière venait d'en bas (Mamassian & Goutcher, 2001; Sun & Perona, 1998). Ces a priori, qui risquent d'être encodés dans les poids de nos connexions synaptiques, sont probablement appris sur de très longs laps de temps ou même transmis héréditairement.

D'autres attentes sont formées sur la base de probabilités conditionnelles. Par exemple, le contexte dans lequel on se trouve va nous renseigner sur les objets les plus probables d'apparaître : on s'attend à percevoir un séchoir dans la salle de bain et une baguette à la boulangerie, mais pas vraiment une baguette dans la salle de bain et un séchoir à la boulangerie. Alors que la plupart de ces associations sont apprises tout au long de la vie, leur influence dépend d'un stimulus ou d'un événement récent (e.g., entrer, ou se diriger vers, la boulangerie). La modulation de la perception par celles-ci doit donc se baser sur un mécanisme plus rapide que l'encodage dans les poids synaptiques. Cette modulation s'effectue probablement par traitement descendant : les régions de plus haut niveau communiqueraient les attentes aux régions de plus bas niveau qui moduleraient le traitement de l'information sensorielle reçue en conséquence (voir section 1.4.3).

Une interaction entre les a priori à long terme encodés dans nos connexions synaptiques et les attentes à court terme basées sur des probabilités conditionnelles est également possible. En effet, nos a priori à long terme peuvent être modulés par des expériences plus récentes, du moins à la suite d'un entraînement prolongé (Adams, Graf & Ernst, 2004; Sotiropoulos, Seitz & Seriès, 2011).

1.4.3 Mécanismes sous-tendant l'effet des attentes

Nous allons désormais nous concentrer sur les attentes à court terme basées sur des probabilités conditionnelles qui sont initiées à la suite d'un stimulus. Plusieurs modèles ont été proposés pour expliquer leur influence, le plus notable d'entre eux étant le principe du cerveau prédictif (Feldman & Friston, 2010; Friston, 2005, 2010). Selon cette théorie, le cerveau serait constamment en train de prédire son environnement. Des prédictions seraient transmises des aires de plus haut niveau aux aires de plus bas niveau adjacentes, et seule l'information sensorielle ne correspondant pas à ces prédictions serait transmise de manière ascendante (Feldman & Friston, 2010; Friston, 2005; voir également Mumford, 1992; Ullman, 1995; Rao & Ballard, 1999). Cette théorie ne semble cependant pas expliquer l'influence d'attentes basées sur la cooccurrence de stimuli arbitraires (e.g., s'attendre à voir un couteau dans une cuisine; Hindy, Ng & Turk-Browne, 2016) : ces attentes, qui peuvent influencer le traitement dans les mêmes régions visuelles, requièrent un

mécanisme autre que la transmission d'information des aires adjacentes. Une possibilité est la complétion de patterns dans l'hippocampe, où l'exposition à une partie d'un stimulus réactive la représentation du stimulus entier (Hindy et al., 2016; Leutgeb & Leutgeb, 2007; Marr, 1971). D'autres auteurs rapportent des prédictions provenant spécifiquement du cortex orbitofrontal (Bar, 2004; Bar et al., 2006), du striatum ventral (O'Doherty et al., 2004) et du cortex préfrontal (Summerfield et al., 2006; Summerfield & Koechlin, 2008).

Les attentes peuvent agir sur notre perception de deux façons distinctes. D'abord, elles peuvent moduler le signal sensoriel en tant que tel. Selon cette perspective, qui est compatible avec les théories présentées ci-haut, le traitement d'un stimulus dans les aires sensorielles diffère en présence d'attentes. Cette perspective est supportée par plusieurs résultats empiriques selon lesquels les patterns d'activité dans les régions visuelles de bas niveau reflètent de manière spécifique les stimuli attendus, et ce, même si ceux-ci ne sont pas présentés (Figure 1.6; Hindy et al., 2016; Kok, Rahnev, Jehee, Lau & de Lange, 2012; Kok, Failing & de Lange, 2014; Kok, Mostert & de Lange, 2017). En utilisant la méthode de la corrélation inverse (voir section 1.5.2), certaines études ont également observé que les attentes biaisaient les représentations internes vers les stimuli attendus, comparativement à lorsqu'il n'y avait pas d'attentes (Cheadle, Egner, Wyart, Wu & Summerfield, 2015; Wyart et al., 2012a).

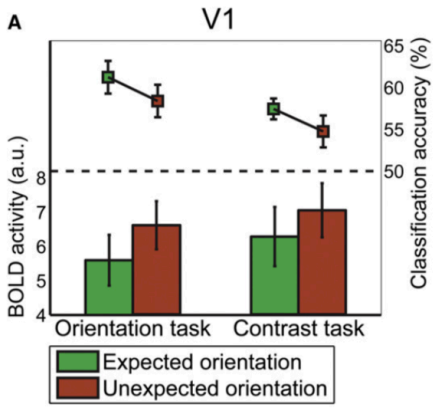


Figure 1.6. Dans cette étude en IRMf, les auteurs démontrent que l'orientation de la grille sinusoïdale présentée peut être mieux décodée si cette grille est attendue, malgré une réponse BOLD globalement réduite. Figure adaptée de Kok et al. (2012).

Une manière alternative selon laquelle les attentes peuvent moduler notre perception est par une modulation du critère décisionnel. Selon cette vue, compatible avec la théorie de détection de signaux et la théorie de décision Bayésienne, le traitement du stimulus dans les régions sensorielles demeure le même, mais c'est les régions de plus haut niveau chargées de la décision qui changeraient leur réponse selon l'attente (Bang & Rahnev, 2017). Cette vue est soutenue par quelques études. Par exemple, Bang et Rahnev (2017) utilisent des indices (pour initier une attente) qui sont soit présentés avant le

stimulus ou après celui-ci, et les sujets doivent ensuite effectuer une tâche de catégorisation. En effectuant des analyses de corrélation inverse (voir section 1.5.2), ils observent des patterns similaires d'utilisation de l'information entre les deux conditions (indices avant vs après), ce qui suggère que les attentes ne modulent pas le traitement sensoriel dans leur design expérimental. Dans une autre étude, Rungratsameetaweemana, Itthipuripat, Salazar & Serences (2018) démontrent quant à eux que des composantes EEG typiquement associées au traitement spécifiquement sensoriel ne sont pas affectées par la présence d'attentes (malgré leur sensibilité à la quantité d'information disponible), mais que des signaux reflétant l'ensemble du traitement ou le traitement de plus haut niveau le sont.

Le débat entre ces deux vues opposées est loin d'être réglé : des études futures seront nécessaires afin de résoudre la question. Cependant, nous pouvons tout de même explorer les conséquences et les mécanismes derrière l'effet des attentes, même si leur caractère sensoriel ou décisionnel n'est pas complètement élucidé.

1.4.4 Attentes et routines visuelles

Comme mentionné ci-haut, quelques études ont utilisé à ce jour la technique de la corrélation inverse pour examiner le contenu des représentations internes (Bang & Rahnev, 2017; Cheadle et al., 2015; Wyart et al., 2012a; voir section 1.5.2). Ces études, qui ont utilisé des grilles sinusoïdales comme stimuli, ont découvert une modulation de la sensibilité à différentes orientations par les attentes.

L'utilisation de ces stimuli très simples, en plus d'être peu écologique, nous empêche de poser une question d'importance : comment l'attente d'un objet spécifique module-t-elle l'utilisation de ses attributs pour le reconnaître? En effet, en utilisant des stimuli composés d'une unique composante (i.e. une grille sinusoïdale d'une orientation et fréquence spatiale précises), la composante et l'objet sont identiques : il devient alors impossible d'examiner comment l'attente d'un objet module l'utilisation de différents attributs. Cependant, les objets de la vie de tous les jours sont composés de multiples attributs auxquels sont sensibles différentes parties de notre système visuel et l'attente d'un

objet pourrait moduler l'utilisation de ces attributs différemment. Il est envisageable que seule une partie des attributs d'un objet soit utilisée pour confirmer la présence d'un objet auquel on s'attend, alors que l'ensemble des attributs représentés en mémoire risque d'être utilisé si aucune attente spécifique n'est présente.

Une autre question d'intérêt concerne les moments auxquels les attentes modulent l'utilisation d'information durant la fixation de l'objet. En effet, différents attributs sont typiquement utilisés à différents moments pour catégoriser un objet ou un visage (e.g., Caplette et al., 2016; Dupuis-Roy et al., 2019; Vinette, Gosselin & Schyns, 2004); ainsi, les attentes risquent de moduler l'utilisation de différents attributs à différents moments. De plus, la présence d'une attente pourrait s'apparenter conceptuellement à un traitement partiel et les premières étapes d'un échantillonnage séquentiel d'information pourraient ainsi être accélérées ou éliminées complètement.

Les fréquences spatiales (FS) ne sont pas utilisées de manière uniforme à travers le temps durant la fixation d'un objet (Caplette et al., 2016, 2017a; Hughes et al., 1996; Schyns & Oliva, 1994). Durant une tâche de reconnaissance d'objets avec des objets de la vie de tous les jours, il a été rapporté que les sujets débutent typiquement à utiliser les plus basses FS qu'ils continuent d'échantillonner durant toute la fixation, puis ils utilisent les plus hautes FS à la fin de la fixation seulement (Caplette et al., 2016, 2017a). Cet échantillonnage précoce des basses FS, combiné à leur rapide traitement dans le cerveau (Bullier & Nowak, 1995; Mazer et al., 2002), pourrait permettre aux sujets de rapidement se former une hypothèse à propos de l'identité de l'objet; cette hypothèse pourrait ensuite moduler l'échantillonnage des plus hautes FS (Bar, 2003; Bar et al., 2006; voir section 1.3.1.2). Ainsi, si une hypothèse est déjà présente, sous la forme d'une attente, avant la perception de l'objet, et que les attentes s'apparentent à des hypothèses formées par le traitement des basses FS, il est possible que l'utilisation des basses FS soit réduite. Cette utilisation pourrait être réduite principalement au début de la fixation, au moment où l'information échantillonnée peut potentiellement influencer l'échantillonnage à la fin de la fixation. Il est également envisageable que d'autres FS soient moins utilisées tout au long de la fixation parce que l'information qu'elles contiennent est maintenant apportée par l'attente. Si l'hypothèse s'avère invalide, l'utilisation d'information pourrait changer au moment où l'invalidité de la prédiction est réalisée par le système visuel : alors que la

sensibilité aux attributs représentés de l'objet attendu était probablement initialement augmentée, le système visuel risque d'opter pour une stratégie plus générique si l'attente est invalide.

1.5 Considérations méthodologiques

1.5.1 La question de l'information en catégorisation visuelle

Gosselin & Schyns (2002) introduisirent un cadre théorique simplifiant la question de l'information en catégorisation visuelle. Sa formulation est la suivante: $R \otimes A \approx P$, où R (*representation*) correspond à la représentation visuelle interne, A (*available*) correspond à l'information disponible pour effectuer une tâche donnée et P (*potent*) correspond à l'information utilisée par l'observateur pour faire la tâche; l'opérateur $\otimes$ symbolise une interaction et l'opérateur $\approx$ symbolise une similarité (Figure 1.7; Dupuis-Roy, 2014; Gosselin & Schyns, 2002). L'information A est une quantité statistique, qui peut être retrouvée par une analyse mathématique de l'ensemble des stimuli (en supposant certaines contraintes). Un observateur idéal possédant des ressources perceptives et computationnelles illimitées utiliserait l'ensemble de cette information pour effectuer une tâche. Dans un tel cas, $R = A = P$ et le problème de la catégorisation visuelle est trivial. Les observateurs humains sont toutefois limités de toutes sortes de manières et R ne contiendra pas toutes les informations de A. De plus, R pourrait contenir des informations que A ne contient pas. En effet, R, qui correspond à la représentation d'un objet, d'un concept ou d'une catégorie qu'un observateur a internalisé au cours de sa vie, pourrait contenir des informations qui ne sont pas disponibles dans un stimulus présent à un moment précis. Cette représentation pourrait même contenir des informations qui ne sont pas utiles pour catégoriser un stimulus dans la moyenne des situations. De telles informations pourraient avoir été acquises à la suite d'expositions à des situations non-représentatives. Évidemment, c'est peu discuté, mais R risque de changer à la suite de, ou même durant, une tâche de catégorisation. En ce sens, et selon une perspective Bayésienne, R pourrait être considéré comme un a priori (*prior*) influençant la perception et étant sans cesse mis à

jour selon les expériences récentes. Nous supposerons cependant ici une représentation qui est fixe pour la durée de la tâche. Finalement, *P* correspond à l'information utilisée par l'observateur : il s'agit de l'information qui est à la fois disponible et représentée. Dans le cas d'un observateur linéaire qui effectue une tâche en vérifiant simplement si l'information présentée correspond à sa représentation, *P* correspond à une simple intersection entre *A* et *R* et pourrait être retrouvé en estimant ces quantités (de même que *R* pourrait être calculé en estimant *A* et *P*) et en effectuant une division de Hadamard (élément par élément); cependant, si l'observateur applique une stratégie non-linéaire (ce qui est probable de survenir dans des tâches de catégorisation de haut niveau), *P* ne peut pas simplement être retrouvé l'aide de *A* et *R* (Murray & Gold, 2004a, 2004b; Murray, 2011; Gosselin & Schyns, 2004).

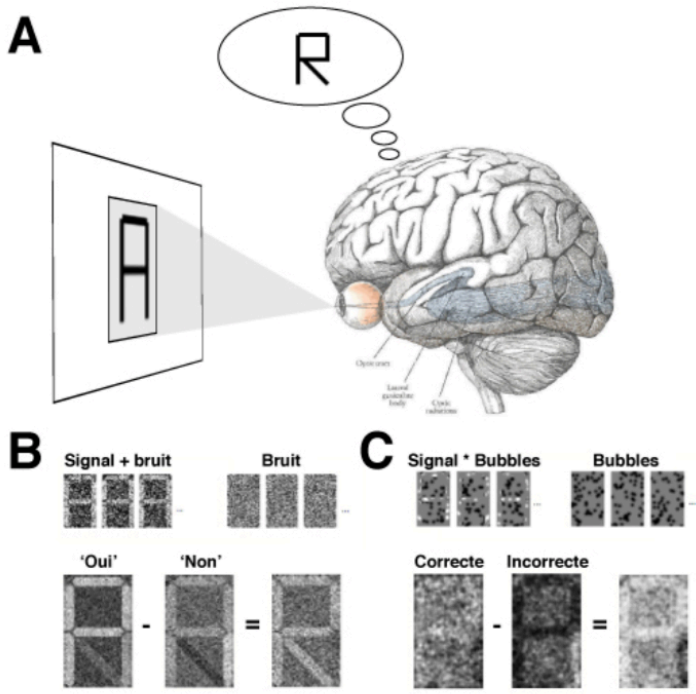


Figure 1.7. Illustration du cadre théorique *RAP*. A) La représentation par le cerveau de l'image à détecter consiste en la lettre R (information représentée) alors que l'information disponible dans l'image lors de cet essai est la lettre A (information disponible). B) En utilisant la corrélation inverse classique (voir section 1.5.2.1), il est possible de retrouver l'information représentée (la lettre R). C) En utilisant *Bubbles* (voir section 1.5.2.2), on retrouve l'information utilisée (soit la lettre P, qui consiste, dans ce cas-ci, en l'intersection des lettres A et R). Figure tirée de Dupuis-Roy (2014), après avoir été adaptée de Gosselin & Schyns (2002).

Illustrons maintenant ce que représentent chacun de ces termes dans le contexte d'une tâche de reconnaissance d'objets. Supposons qu'un objet aléatoire parmi un ensemble de cinq objets différents en tons de gris est présenté à un observateur à chaque essai. L'information disponible *A* correspond à toute l'information que l'observateur peut utiliser pour effectuer la tâche : en supposant une stratégie de reconnaissance pixel-par-pixel (peu probable mais plus simple pour des fins d'illustration), il s'agirait des variations dans l'intensité des pixels individuels. Les pixels les moins informatifs seront ceux qui sont identiques d'un objet à l'autre, probablement ceux du fond si les objets sont illustrés sur un fond standard identique. Les pixels les plus informatifs seront ceux dont l'intensité est distribuée de manière régulière à travers les différents stimuli. Entre les deux, certains pixels dont l'intensité permet de distinguer certains stimuli mais pas d'autres seront plus ou moins informatifs. L'information représentée *R* correspond à la représentation interne qu'a l'individu de ce qui permet de distinguer l'ensemble des stimuli. Cette représentation a été internalisée au cours de sa vie avant l'expérience (toujours en supposant que la représentation reste fixe pendant la durée de la tâche), suite à son exposition à ces objets. Par exemple, l'observateur pourrait avoir internalisé uniquement qu'un ensemble de pixels à un certain endroit dans le champ visuel permet de bien distinguer ces cinq objets, même si en réalité d'autres pixels le permettraient, ou peut-être même malgré le fait que les stimuli de l'expérience ne sont pas informatifs à cet endroit. Finalement, l'information utilisée *P* correspondrait, assumant une stratégie linéaire, à l'intersection de *R* et de *A*. Si l'intersection est vide parce que *R* ne contient pas d'information utile, *P* ne contiendra aucune information et l'observateur ne pourra effectuer la tâche.

Évidemment, l'exemple ci-haut repose sur plusieurs postulats simplificateurs. Par exemple, l'observateur n'applique probablement pas une stratégie de reconnaissance pixel par pixel : dans le cas d'une tâche de reconnaissance d'objets, il utilise probablement des attributs de plus haut niveau (plus abstraits) basés sur des combinaisons de pixels. En d'autres mots, l'espace génératif utilisé par les sujets ne risque pas d'être l'espace photométrique, i.e. le plan de l'image. La question de l'espace génératif (ou espace de recherche) utilisé par les sujets est une question fondamentale en reconnaissance d'objets et est le sujet d'intenses recherches (voir section 1.2). Même si les régions de plus bas niveau (V1, V2) sont relativement bien modélisées depuis de nombreuses années, le format

des représentations dans les régions de plus haut niveau (V4, IT, etc.) demeurait jusqu'à tout récemment totalement mystérieux. Ces dernières années, plusieurs études ont observé des similarités entre les représentations dans les régions cérébrales de haut niveau et les représentations dans les dernières couches des réseaux de neurones artificiels profonds (DNN; Bashivan et al., 2019; Khaligh-Razavi & Kriegeskorte, 2014; Kriegeskorte, 2015; Yamins & DiCarlo, 2016; Yamins et al., 2014) : ces représentations semblent pour l'instant notre meilleur modèle de l'espace génératif employé par les sujets.

1.5.2 Méthodes d'images de classification

Nous allons maintenant nous attarder à certaines méthodes qui sont couramment utilisées pour révéler les informations discutées ci-haut. Ces méthodes peuvent être incluses dans une même famille de méthodes d'images de classification (voir Murray, 2011, pour une revue de la littérature).

1.5.2.1 La corrélation inverse

Ahumada (1996) introduisit l'outil méthodologique qu'est la corrélation inverse (*reverse correlation*) à la psychophysique visuelle. Dans une expérience typique de corrélation inverse, le stimulus à chaque essai consiste en un de deux signaux (ou un signal vs l'absence de signal) présenté à faible contraste auquel est additionné une plage de bruit blanc Gaussien (différent à chaque essai). Le sujet tente ensuite d'identifier lequel des deux signaux était présent et indique sa réponse (par exemple, en appuyant sur une touche d'un clavier). Les plages de bruit présentées pendant l'expérience sont ensuite analysées. Typiquement, les plages de bruit menant à une réponse (peu importe le stimulus présenté) sont moyennées et soustraites des plages de bruit menant à l'autre réponse : le résultat est appelé « image de classification » et permet de révéler les représentations internes (*R*; voir section 1.5.1) des sujets (Figure 1.7b; Ahumada, 1996). La logique derrière cette méthode est la suivante. Par hasard, lors de certains essais, le bruit sera plus élevé dans certaines régions que dans d'autres. Ainsi, par hasard, le sujet risque d'identifier un signal (alors qu'il n'est pas nécessairement présenté) plus dans certains essais que dans d'autres. La

moyenne des champs de bruit de tous les essais auxquels le sujet a indiqué avoir identifié ledit signal révélera donc ce qui a mené le sujet à indiquer cette réponse, i.e. sa représentation interne du stimulus. Une autre manière de concevoir une image de classification est comme une image de corrélations, la valeur de chaque pixel indiquant à quel point la luminance de ce pixel corrèle avec la réponse du sujet (Murray, 2011). La corrélation inverse peut également être employée avec une réponse neurophysiologique (telle que le taux de déclenchement d'un neurone; voir Ringach & Shapley, 2004, pour une revue de la littérature) plutôt qu'une réponse comportementale.

Gosselin & Schyns (2003) ont poussé cette idée à la limite en ne présentant à chaque essai que des plages de bruit, sans signal. Les sujets, cependant, avaient été avertis qu'un stimulus était présent derrière le bruit lors de 50% des essais. Ils devaient donc indiquer à chaque essai s'ils pensaient avoir perçu le stimulus ou non. De manière similaire à la corrélation inverse classique, les plages de bruit menant à chaque réponse sont ensuite moyennées entre elles puis les plages menant à la réponse « stimulus absent » sont soustraites de celles menant à la réponse « stimulus présent ». Dans un premier temps, les auteurs ont pu révéler la représentation interne de la lettre « S » de trois observateurs (alors qu'un tel stimulus n'avait jamais été présenté au cours de l'expérience). Dans un second temps, les auteurs ont pu révéler la représentation interne d'un sourire (il y avait dans cette seconde expérience le contour du visage qui était partiellement présenté, mais aucune information n'était présente dans la région de la bouche; Gosselin & Schyns, 2003). Cette méthode a également été utilisée par la suite avec d'autres stimuli (e.g., Gosselin, Bacon & Mamassian, 2004; Jack, Caldara & Schyns, 2012; Mangini & Biederman, 2004; Morin-Duchesne, Gosselin, Fiset & Dupuis-Roy, 2014) et d'autres variables dépendantes que l'exactitude de la réponse comportementale (spécifiquement, l'activité EEG; Smith, Gosselin & Schyns, 2012). Notons également que la logique de la corrélation inverse classique, avec ou sans signal, peut également être appliquée à d'autres espaces de recherche que l'espace photométrique.

1.5.2.2 Bubbles

Les méthodes discutées ci-haut permettent de révéler les représentations internes des sujets : elles permettent ainsi de révéler des attributs représentés mais jamais présentés dans le stimulus. Alternativement, il peut parfois être plus pertinent de retrouver l'information utilisée pour effectuer une tâche (*P*; voir section 1.5.1). Gosselin & Schyns (2001) introduisirent une méthode qu'ils nommèrent *Bubbles* pour pallier ce manque. Nous la décrivons de manière détaillée ci-bas.

Dans une expérience typique utilisant *Bubbles*, le stimulus à chaque essai consiste en une image dissimulée par un masque de « bulles » en révélant certaines parties seulement, de manière aléatoire (différemment à chaque essai). Plus précisément, à chaque essai, une matrice de la taille des images composée de zéros avec quelques « uns » dans des positions aléatoires est créée. Notons que le nombre de « uns » est soit gardé constant, soit ajusté au cours de l'expérience pour maintenir la performance du sujet au même niveau. Plutôt que d'ajuster strictement le *nombre* de « uns », il est en fait recommandé (afin d'éviter des dépendances entre les pixels) d'ajuster la probabilité que chaque pixel soit un « un »; le bruit utilisé se réduit ainsi à un bruit blanc de Bernoulli. La matrice de bruit est ensuite lissée par une Gaussienne 2D (l'écart-type de cette Gaussienne est normalement déterminé avant l'expérience) : cette matrice constitue le masque de bulles. Puis, un produit de Hadamard (élément par élément) est effectué entre ce masque (seuillé pour que toutes les valeurs se retrouvent entre 0 et 1) et l'image à présenter lors de l'essai : en résulte alors un stimulus où certaines parties de l'image seulement sont révélées. Ainsi, la méthode *Bubbles* utilise du bruit multiplicatif alors que la corrélation inverse classique utilise du bruit additif : il s'agit d'une différence fondamentale entre les deux méthodes qui fait en sorte que celles-ci ne retrouvent pas le même type d'information. Le sujet doit ensuite indiquer sa réponse, typiquement la catégorie à laquelle le stimulus appartient selon lui. Les masques de bulles présentés durant l'expérience sont finalement analysés. Les masques de bulles menant à une réponse incorrecte peuvent être soustraits des masques menant à une réponse incorrecte afin d'obtenir une image de classification (notons ainsi qu'ici, l'exactitude des réponses est utilisée plutôt que la réponse elle-même). Cette image de classification représentera l'information utilisée pour faire la tâche (Figure 1.7c). Cette analyse est un cas particulier de la somme pondérée, qui est également adaptée à des

réponses continues plutôt que binaires. La somme pondérée est également équivalente à une régression linéaire multiple dans ce cas puisque la matrice de covariance dans une telle expérience est une matrice identité multipliée par un scalaire (étant donné le caractère aléatoire de l'échantillonnage; si la matrice avant lissage Gaussien est utilisée); l'utilisation d'une régression serait cependant impossible en pratique étant donné le nombre de paramètres à estimer typiquement plus grand que le nombre d'essais et l'impossibilité d'inverser la matrice de covariance – à moins d'utiliser une forme de régularisation (voir les articles 2 et 3 de cette thèse). L'information mutuelle peut également être utilisée afin de retrouver des relations qui ne sont pas nécessairement linéaires (Ince et al., 2015, 2016, 2017; Schyns, Thut & Gross, 2011; Zhan, Ince, van Rijsbergen & Schyns, 2019). Les masques de bulles sont ensuite fréquemment transformés en scores *Z* ou *t* à l'aide d'une distribution nulle estimée empiriquement (soit en permutant les essais, soit en prenant une région qu'on sait non-informative) ou en effectuant un test paramétrique (Caplette, McCabe, Blais & Gosselin, 2017b; Chauvin, Worsley, Schyns, Arguin & Gosselin, 2005; Gosselin & Schyns, 2005).

Bien qu'originalement utilisée dans l'espace photométrique (e.g., Dupuis-Roy, Fortin, Fiset & Gosselin, 2009; Gosselin & Schyns, 2001; Rutishauser et al., 2011, 2013), *Bubbles* peut être généralisée à d'autres espaces continus; en retirant la contrainte de lissage Gaussien, la méthode peut même être généralisée à des espaces discontinus. La généralisation à d'autres espaces est cruciale pour investiguer des tâches de haut niveau. En effet, les sujets ne risquent pas en réalité d'utiliser l'espace génératif photométrique. Un tel espace serait particulièrement inapproprié pour investiguer les représentations des objets de la vie de tous les jours, étant donné la variation énorme dans l'intensité des pixels à travers les objets (presque tous les pixels risquent d'être utilisés) et la nature 3D des objets (un même objet pourrait être présenté dans des angles différents et les pixels utilisés changeraient alors complètement). Par exemple, *Bubbles* a été utilisée dans l'espace des FS (e.g., Caplette, West, Gomot, Gosselin & Wicker, 2014; Thurman & Grossman, 2011; Willenbockel et al., 2010), l'espace circulaire des orientations (Duncan et al., 2017), un espace photométrique × bandes de FS (e.g., Adolphs et al., 2005; Adolphs, Spezio, Parlier & Piven, 2008; Gosselin & Schyns, 2001; Schyns, Bonnar & Gosselin, 2002; Spezio, Adolphs, Hurley & Piven, 2007; Tardif et al., 2017), l'espace temporel (Blais et al., 2013;

Jalali, Martin, Murphy, Solomon & Yarrow, 2018; Thurman & Grossman, 2008), un espace photométrique × temps (e.g., Blais et al., 2009; Fiset et al., 2009; Jalali et al., 2018; Vinette et al., 2004; voir également les articles 2 et 3 de cette thèse), un espace FS × temps (Caplette et al., 2016, 2017a; Estephan et al., 2018; voir également l'article 1 de cette thèse), un espace photométrique × bandes de FS × temps (Blais, Roy, Fiset, Arguin & Gosselin, 2012) et en audition dans l'espace fréquences × temps (Mandel, Yoho & Healy, 2016) et dans l'espace modulation fréquentielle × modulation temporelle (Venezia et al., 2016). Un espace de recherche basé sur les réseaux de neurones artificiels profonds (DNN) pourrait également être utilisé dans le futur. Ces espaces sont de différentes dimensionnalités (d'une à quatre dimensions, certaines dimensions étant parfois discontinues) et requerront donc des Gaussiennes de différentes dimensionnalités.

Différentes variables dépendantes peuvent également être utilisées. Mis à part l'exactitude des réponse comportementales (e.g., Adolphs et al., 2005; Fiset et al., 2008; Gibson, Lazareva, Gosselin, Schyns & Wasserman, 2007; Gosselin & Schyns, 2001; Langner, Beckner & Rinck, 2009; Tadros, Dupuis-Roy, Fiset, Arguin & Gosselin, 2013; Willenbockel et al., 2010; voir également l'article 1 de cette thèse) et les temps de réponse (e.g., Caplette et al., 2014; Dupuis-Roy et al., 2009; Schyns et al., 2002; Willenbockel et al., 2010), des études passées ont également utilisé des mesures de neuroimagerie. L'activité électroencéphalographique (EEG; e.g., Smith et al., 2006; Smith, Gosselin & Schyns, 2007; voir également l'article 2 de cette thèse), l'activité magnétoencéphalographique (MEG; Smith et al., 2009; Zhan et al., 2019; voir également l'article 3 de cette thèse), l'imagerie par résonance magnétique fonctionnelle (IRMf; Smith et al., 2008) et le taux de déclenchement de neurones unitaires (e.g., Wang et al., 2014; Rutishauser et al., 2013) sont parmi les mesures utilisées. Dans tous les cas, une mesure continue d'amplitude peut être utilisée, ce qui augmente la richesse des données par rapport aux mesures comportementales, ou une division en catégories (*binning*) peut être effectuée. L'utilisation de la neuroimagerie multiplie évidemment la dimensionnalité des données puisque maintenant, ce n'est pas seulement une image de classification qui est obtenue, mais une image de classification par électrode, par source, par voxel ou par neurone, et potentiellement également par latence. Évidemment, une réduction de la dimensionnalité peut être effectuée par un choix de régions d'intérêt, un moyennage ou une analyse par

composantes principales. Il est également possible d'utiliser simultanément la neuroimagerie et le comportement, afin de retrouver à la fois l'information qui est représentée uniquement dans le cerveau et pas dans le comportement, l'information qui est représentée uniquement dans le comportement et pas dans notre mesure de l'activité cérébrale (évidemment, le cerveau représente certainement l'information si elle mène à un comportement, alors il s'agit ici d'information que notre méthode de neuroimagerie ne permet pas d'accéder) et l'information qui est représentée à la fois dans le cerveau et dans le comportement (Ince et al., 2017; Schyns & Ince, soumis). La mesure utilisée pour obtenir ces informations est l'information interactive (*interaction information*), qui correspond essentiellement à une information mutuelle pour plus de deux variables (McGill, 1954); une analyse de variance commune (*commonality analysis*; Hebart, Bankson, Harel, Baker & Cichy, 2018; Seibold & McPhee, 1979) pourrait également être effectuée si l'on souhaite rester dans le domaine linéaire en utilisant la variance comme quantité d'intérêt plutôt que l'information au sens de la théorie de l'information. Ces méthodes pourraient aussi être utilisées pour vérifier la variance (ou l'information) commune à d'autres variables (par exemple l'activité cérébrale à deux différents moments ou à deux différentes électrodes; Ince et al., 2016, 2017).

En appliquant le cadre d'analyse *RAP*, nous pouvons déduire que *Bubbles* en neuroimagerie permet de retrouver l'information utilisée par une région cérébrale spécifique (plutôt que par le sujet, ou le cerveau entier), pour effectuer une tâche donnée. Similairement, la corrélation inverse classique permettrait de retrouver l'information représentée par une région cérébrale spécifique. Cependant, il faut mettre l'accent sur le fait que l'information utilisée par une région cérébrale ne sera pas nécessairement utilisée comportementalement.

Dans cette thèse, nous utiliserons *Bubbles* plutôt que la corrélation inverse puisque nous nous intéressons davantage à l'information utilisée qu'à l'ensemble de l'information représentée. Nous souhaitons avant tout évaluer le traitement de l'information qui est disponible dans l'input sensoriel, particulièrement en neuroimagerie. De plus, *Bubbles* semble généralement être une technique plus puissante lorsque des tâches de catégorisation de haut niveau sont employées (un beaucoup plus grand nombre d'essais est nécessaire en utilisant la corrélation inverse; e.g., Jack et al., 2012; Nestor & Tarr, 2008a).

1.5.3 Corrélation inverse temporelle

Comme brièvement mentionné ci-haut, la dimension temporelle peut faire partie de l'espace de recherche. L'utilisation de cette dimension implique que le traitement de l'information selon son moment de réception sur la rétine est investigué (même si ce fait n'est parfois pas apprécié et que les résultats sont interprétés en termes de traitement temporel). Quelques études ont investigué cette question de manière comportementale, soit avec la corrélation inverse classique (Nagai et al., 2007; Neri & Heeger, 2002; Neri & Levi, 2007, 2008) ou avec *Bubbles* (Bais et al., 2009, 2012; Caplette et al., 2016, 2017a; Dupuis-Roy et al., 2019; Fiset et al., 2009; Vinette et al., 2004). Spécifiquement, les études utilisant la corrélation inverse classique présentent à chaque essai un stimulus avec un signal constant (Neri & Levi, 2007; Nagai et al., 2007) ou dynamique (Neri & Heeger, 2002; Neri & Levi, 2008) et y ajoutent un bruit dynamique. Ces études se sont concentrées surtout sur l'étude de mécanismes de bas niveau (tels la détection) en utilisant des stimuli très simples et une faible résolution (potentiellement suite à l'absence de lissage et à la puissance statistique réduite). Les études utilisant *Bubbles* présentent quant à elles à chaque essai un signal constant (Blais et al., 2009; Caplette et al., 2016, 2017a; Dupuis-Roy et al., sous presse) ou dynamique (Blais et al., 2012) multiplié (élément par élément) par un bruit dynamique lissé temporellement. Ces études portent davantage sur la reconnaissance d'objets et de visages. De manière intéressante, toutes ces études ont permis de révéler des stratégies d'utilisation de l'information stables à travers les individus.

En neuroimagerie, certaines études ont utilisé un échantillonnage temporel aléatoire (typiquement, en variant la luminance d'un stimulus ou de l'écran au complet de manière aléatoire à travers le temps) avec une réponse neurale temporelle (spécifiquement, en EEG; Lalor, Pearlmutter, Reilly, McDarbt & Foxe, 2006; VanRullen & MacDonald, 2012; voir également Crosse, Di Liberto, Bednasr & Lalor, 2016; Smith & Kutas, 2015). Cependant, ces études effectuaient une analyse de corrélation croisée (*cross-correlation*) et non une analyse de corrélation inverse, ce qui revenait essentiellement à moyenner ensemble les réponses à l'information reçue à tous les moments (pour une revue des deux méthodes, voir

Lalor, Pearlmutter & Foxe, 2009). C'était dans leur cas la seule analyse sensée, puisqu'il n'y avait pas d'essais à proprement dit et que les stimuli duraient de nombreuses secondes.

Notons que toutes ces méthodes supposent que le cerveau est un système linéaire invariable au temps (*Linear Time-Invariant*, ou *LTI*; voir Lalor et al., 2009), c'est-à-dire que sa réponse à un événement instantané (*Impulse Response Function*, ou *IRF*) est toujours la même et que sa réponse à une somme d'événements se succédant rapidement est simplement la somme de ces IRF. Ce postulat est évidemment faux, mais permet d'obtenir des approximations utiles (Lalor et al., 2009). De plus, l'utilisation de la corrélation inverse ou de la corrélation croisée comporte des avantages par rapport à d'autres méthodes telles que la méthode des potentiels évoqués, notamment l'absence de stimuli débutant abruptement (on peut exclure des analyses cette partie des stimuli; VanRullen & Macdonald, 2012) et la puissance statistique plus élevée (Lalor et al., 2006). Finalement, il est toujours possible avec de telles méthodes d'utiliser des analyses plus sophistiquées pour investiguer la réponse à des combinaisons d'événements (e.g., la réponse à la somme de deux événements se suivant à 10 ms d'intervalle; Lalor et al., 2009).

1.5.4 Une méthode pour décomposer le traitement selon le moment d'échantillonnage

En neuroimagerie, plusieurs études ont utilisé des stimuli temporels et varié certains paramètres afin d'investiguer leur effet sur l'activité cérébrale. Des études ont vérifié l'effet de l'intervalle inter-stimuli (e.g., Bacon-Macé, Macé, Fabre-Thorpe & Thorpe, 2005), de la durée des stimuli (e.g., Brisson & Jolicoeur, 2007; Tanskanen, Näsänen, Ojanpää & Hari, 2007) et de l'ordre de présentation de l'information (e.g., basses à hautes FS vs hautes à basses FS; Kauffmann, Chauvin, Pichat & Peyrin, 2015). Aucune étude n'a cependant examiné comment l'information reçue sur la rétine à différents moments est traitée par le cerveau en utilisant la corrélation inverse temporelle, que ce soit avec une réponse neuronale statique ou temporelle.

On peut appliquer la méthode *Bubbles* sur la dimension temporelle, en utilisant un signal neural plutôt que comportemental comme réponse, pour parvenir à ce but. En utilisant un signal de neuroimagerie uniquement spatial (ou du moins que nous

modéliserions pour obtenir un seul estimé par condition par coordonnée spatiale; e.g., IRMf), nous obtiendrions une image de classification temporelle[4] (ou spatiotemporelle si les dimensions spatiales sont investiguées en plus de la dimension temporelle) par coordonnée spatiale (e.g., voxel) : cette image, ou vidéo, nous renseignerait sur le traitement d'une information reçue à un moment spécifique à cet endroit dans le cerveau[5]. En utilisant un signal de neuroimagerie spatiotemporel (e.g., EEG ou MEG), nous obtiendrions une image de classification temporelle ou spatiotemporelle pour chaque coordonnée spatiale (e.g., électrode ou source) et pour chaque moment (latence). De telles images nous indiqueraient comment est traitée, à travers le temps et à travers le cerveau, l'information reçue à un moment spécifique sur la rétine pendant la fixation d'un objet.

Nous arguons qu'il est particulièrement intéressant, malgré la dimensionnalité élevée des données, d'utiliser *Bubbles* dans la dimension temporelle en conjonction avec une autre dimension (e.g., espace photométrique × temps, i.e. *Bubbles* spatiotemporelle), avec une réponse neurale temporelle (ou plus communément spatiotemporelle). Repassons d'abord sur ce que ça implique, dans le cas de *Bubbles* spatiotemporelle. À chaque essai, des informations spatiales spécifiques seraient révélées à différents moments pendant un court laps de temps (normalement 200 ms pour rester globalement à l'intérieur d'une fixation oculaire). Typiquement, cet échantillonnage serait lisse dans les dimensions spatiales et temporelles : une matrice aléatoire 3D espace × temps serait créée et lissée par une Gaussienne 3D (e.g., Vinette et al., 2004). L'activité électrique ou magnétique du scalp du sujet serait enregistrée pendant ce temps (probablement en conjonction avec une réponse comportementale). Une somme pondérée ou une régression régularisée peut ensuite être effectuée entre les matrices d'échantillonnage et l'activité cérébrale à travers les essais.

[4] Notons que les latences de cette image, ou vidéo, de classification peuvent être quelque peu arbitraires. En effet, elles dépendent de la durée de présentation choisie pour l'expérience (voir section 5.5.1.1).

[5] Évidemment, une telle méthode moyenne à travers plusieurs essais et donc repose sur le postulat implicite que ces essais sont comparables en termes d'échantillonnage de l'information. Pour remédier à ce potentiel problème, nous pouvons nous assurer que la phase des oscillations soit réinitialisée à chaque essai et/ou nous assurer que les sujets peuvent prédire le début du prochain stimulus (par exemple en gardant l'intervalle inter-stimuli ou l'intervalle réponse-stimulus constant).

Les images de classification obtenues dans une telle expérience pourraient être réarrangées pour créer des cartes temps (de présentation, ou réception sur la rétine) × temps (de traitement, depuis le début du stimulus) pour chaque source ou électrode. Ces cartes nous permettraient d'observer le traitement temporel du stimulus selon le moment d'arrivée de l'information sur la rétine et ainsi d'ouvrir l'accès à une nouvelle dimension du traitement visuel. Avoir accès à cette information nous permettrait de visualiser directement l'accumulation et l'intégration d'informations reçues à différents moments en examinant le traitement ou le transfert, à un même moment, d'informations reçues à différents moments (voir section 1.3.1.3). On pourrait également visualiser les oscillations perceptuelles ou attentionnelles et les dissocier des oscillations dans le traitement. Il serait aussi possible de visualiser l'effet de routines visuelles séquentielles et d'observer l'effet d'une routine adaptative (voir section 1.3.1.2).

1.6 Objectifs généraux et présentation des articles

La présente thèse comporte trois articles principaux, présentés dans cette section. Nous avons également inclus en annexe un 4e article effectué pendant la même période et partiellement relié aux thèmes de cette thèse. L'auteur de cette thèse est le premier auteur de tous ces articles.

Les trois articles principaux de cette thèse portent sur l'utilisation (ou l'échantillonnage) d'information à différents moments tout au long de la fixation. En ce sens, les objectifs généraux de la thèse sont d'évaluer les différences dans cet échantillonnage à travers le temps, d'évaluer les mécanismes sous-tendant ces différences et de déterminer par quels facteurs ces différences sont modulées.

1.6.1 Premier article

Le premier article de cette thèse évalue l'utilisation des FS à travers le temps lors de la reconnaissance d'objets et sa modulation par l'attente d'un objet spécifique. En effet, on peut s'attendre à ce qu'une stratégie générique d'utilisation de l'information soit effectuée

si le sujet ne s'attend à aucun objet spécifique, alors que si le sujet s'attend à voir un objet spécifique, sa stratégie soit biaisée vers l'utilisation des attributs de l'objet déjà représentés en mémoire. Pour ce faire, nous avons utilisé *Bubbles* dans les dimensions des FS et du temps (bulles 2D). Nous avons révélé une utilisation plus grande des basses FS au début de la fixation lorsque les sujets s'attendaient à voir un objet spécifique. Cette plus grande utilisation peut être expliquée par le fait que les sujets se concentrent maintenant sur des attributs spécifiques de l'objet attendu. De plus, l'utilisation des basses FS plus tard dans la fixation variait significativement selon l'objet spécifique attendu, probablement parce que les attributs qui sont utilisés à ce moment-là varient en termes de FS selon l'objet attendu. Finalement, l'utilisation de hautes FS à la fin de la fixation corrélait avec l'habileté générale des sujets seulement lorsqu'il n'y avait pas d'attentes, indiquant que les meilleurs sujets n'avaient plus besoin d'échantillonner les hautes FS qu'ils échantillonnaient normalement lorsqu'un objet spécifique était attendu.

Cet article a été écrit par Laurent Caplette (LC), Frédéric Gosselin (FG) et Gregory West (GW). LC, FG et GW ont conçu le design expérimental; LC a programmé l'expérience; GW s'est occupé de la collecte des données; LC a analysé les données; LC a écrit le premier jet du manuscrit; LC, FG et GW ont révisé le manuscrit. Cet article a été soumis à un journal scientifique pour publication prochaine; un manuscrit est disponible sur un serveur de prépublications.

1.6.2 Deuxième article

Dans un second article, nous évaluons en EEG le traitement temporel de l'information reçue à différents moments durant la fixation pendant la reconnaissance d'objets. Nous révélons ainsi pour la première fois une nouvelle dimension au traitement visuel. Nous utilisons dans cette étude des visages et des bulles temporelles 1D séparément pour chaque attribut du visage (œil gauche, œil droit, bouche). Nous démontrons que le traitement est significativement différent selon le moment de réception de l'information et ce, à plusieurs électrodes et à plusieurs moments. Nous démontrons de plus que ces différences ne sont pas causées par de simples effets d'amorçage ou d'adaptation, qu'elles sont modulées par

la tâche et donc au moins partiellement d'origine descendante, et qu'elles sont reliées au comportement.

Cet article a été écrit par Laurent Caplette (LC), Robin Ince (RI), Karim Jerbi (KJ) et Frédéric Gosselin (FG). LC, KJ et FG ont conçu le design expérimental; LC a programmé l'expérience; LC s'est occupé de la collecte des données; RI a fourni des fonctions pour l'analyse des données; LC a analysé les données; LC a écrit le premier jet du manuscrit; LC, RI, KJ et FG ont révisé le manuscrit. Cet article a été soumis à un journal scientifique pour publication prochaine; un manuscrit est disponible sur un serveur de prépublications.

1.6.3 Troisième article

Dans un troisième article, nous évaluons le traitement temporel d'information reçue à différents moments à l'aide de la MEG. Nous effectuons une reconstruction de sources et révélons ce traitement dans différents régions cérébrales. Nous démontrons que l'information est traitée de manière hautement variable selon le moment de présentation dans de larges portions des lobes occipitaux et pariétaux. Nous démontrons de plus que l'échantillonnage dans de nombreuses régions est rythmique, c'est-à-dire qu'il oscille significativement à des fréquences entre 7 et 30 Hz. Il s'agit de la première démonstration directe d'un échantillonnage rythmique dans le cerveau. Finalement, nous révélons que différents attributs du visage sont échantillonnés à différentes fréquences, dans une instance de multiplexage fréquentiel.

Cet article a été écrit par Laurent Caplette (LC), Karim Jerbi (KJ) et Frédéric Gosselin (FG). LC, KJ et FG ont conçu le design expérimental; LC a programmé l'expérience; LC s'est occupé de la collecte des données; LC a analysé les données; LC a écrit le premier jet du manuscrit; LC, KJ et FG ont révisé le manuscrit. Cet article est en préparation pour une soumission prochaine à un journal scientifique.

Chapitre 2

Article 1

Object expectations alter information use during visual recognition

Laurent Caplette, Frédéric Gosselin & Greg L. West

Department of Psychology, Université de Montréal, Montréal, Qc, Canada

Corresponding author: Laurent Caplette
Email: laurent.caplette@umontreal.ca
Address: Department of Psychology, University of Montreal, C.P. 6128 succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada

2.1 Abstract

Prior expectations influence how we perceive and recognize objects. While recent studies suggest that sensory representations are altered by expectations, how expectations of everyday objects affect representations remains largely unknown. In this study, we used reverse correlation to reveal with high precision how the use of visual information across time is modulated by everyday object expectations in a recognition task. We used spatial frequencies (SFs) as a continuous feature space. We show an earlier use of low SFs when an object is expected and a significantly variable use of low SFs depending on the expected object mid-fixation. Finally, object expectations reduced the use of high SFs toward the end of the fixation, suggesting that an expectation can replace high SF information. Our results illustrate that different object features are affected differently across time during visual recognition and they reveal how rich internal representations are affected by expectations of real-world complex objects.

Keywords: object recognition, expectation, prediction, representation, information use.

2.2 Significance Statement

What we expect to see in the world influences how we perceive it. As suggested by recent evidence, this may be done by altering internal representations. However, how representations of complex everyday objects are altered when we expect them remains unknown. Here, we show that different object features are affected differently by expectations and at different time points during recognition. These results reveal for the first time the mechanisms underlying the influence of real-world object expectations on visual recognition.

2.3 Introduction

Every day, we encounter a myriad of objects and we recognize them effortlessly within milliseconds. Understanding how we do so remains one of the greatest challenges of cognitive neuroscience. Since the seminal work of Helmholtz in the 19th century, prior expectations are believed to be responsible for a key part of this feat. Valid expectations increase our accuracy and our speed at recognizing objects (Pinto, van Gaal, de Lange, Lamme & Seth, 2015; Stein & Peelen, 2015; Wyart, Nobre & Summerfield, 2012) and they can bias our perception of ambiguous or noisy sensory input in specific directions (Kok, Brouwer, van Gerven & de Lange, 2013; Sterzer, Frith & Petrovic, 2008). The mechanisms through which this influence is exerted are not fully understood. One possibility is that sensory processing itself is altered by expectations (Esterman & Yantis, 2010; Kok, Failing & de Lange, 2014; Kok, Jehee & de Lange, 2012; Kok, Mostert & de Lange, 2017; Wyart et al., 2012; but see Bang & Rahnev, 2017). Accordingly, recent findings show that internal representations reflect expected features, even in the absence of an actual stimulus (Cheadle, Egner, Wyart, Wu & Summerfield, 2015; Kok et al., 2014, 2017; Wyart et al., 2012).

Expectations about complex objects seem to involve different mechanisms than expectations about simple features such as gratings of different orientations (Denison, Piazza & Silver, 2011; Denison, Sheynin & Silver, 2016; Kok, Rait & Turk-Browne, submitted). Little is known however about how expectations of complex real-world objects influence the content of representations. Everyday objects are composed of multiple high- and low-level features, and the expectation of an object may activate representations of different features at different moments during the sampling and accumulation of information. For instance, consider the use of low and high spatial frequencies (i.e. coarse and fine visual information).

Spatial frequencies (SFs) are typically not attended uniformly across time during visual recognition (Hughes, Nozawa & Kitterle, 1996; Schyns & Oliva, 1994). In fact, in a recognition task with everyday man-made objects, subjects typically start using low SFs (coarse information) to identify the objects before they start using higher SFs (fine information; Caplette, Wicker & Gosselin, 2016; Caplette, Wicker, Gosselin & West,

2017). This early sampling of low SFs, combined with their faster processing in the brain (Bullier & Nowak, 1995; Mazer, Vinje, McDermott, Schiller & Gallant, 2002), may allow observers to rapidly form a hypothesis about the object's identity which will modulate the later sampling of high SFs (Bar, 2003; Bar et al., 2006; Bullier, 2001). If a hypothesis is already present before the object's perception, it's possible that the first stages of information acquisition are accelerated or skipped altogether, and that low SFs are less used. It's also possible that the use of other SFs would be reduced, because the information they provide would be provided by the expectation. Whether the use of some features is reduced when an object is expected remains an unanswered question.

In the present study, we investigated for the very first time how the expectation of a specific everyday object influences the use of information across time. We used SFs as a feature space, since (i) it is a well-defined continuous space represented explicitly in the early visual system; (ii) objects are probably represented and discriminable in only a subset of all SFs; (iii) SFs are not sampled uniformly across fixation time during object recognition (Caplette et al., 2016); and (iv) the processing of some SFs has been related to the creation of expectations (Bar, 2003).

2.4 Results

On each trial, participants (n = 59) were exposed to one object image (out of 80 possibilities) followed by an object name (no-expectation condition), or to the object name prior to the presentation of the object image (expectation condition; Figure 2.1a). After the stimuli were presented, participants were asked to indicate whether the name matched the object shown. The object name and image matched on 50% of trials; therefore, for a given object name, the object most probable to appear was the object described by the name (50% vs ~0.6% for every other object; Figure 2.1b). On each trial, the object images had their SFs randomly sampled across time; the amount of information (i.e. the number of "Bubbles") presented to the participant was adjusted on a trial-by-trial basis to maintain accuracy around 75% (see Methods; Figure 2.1c).

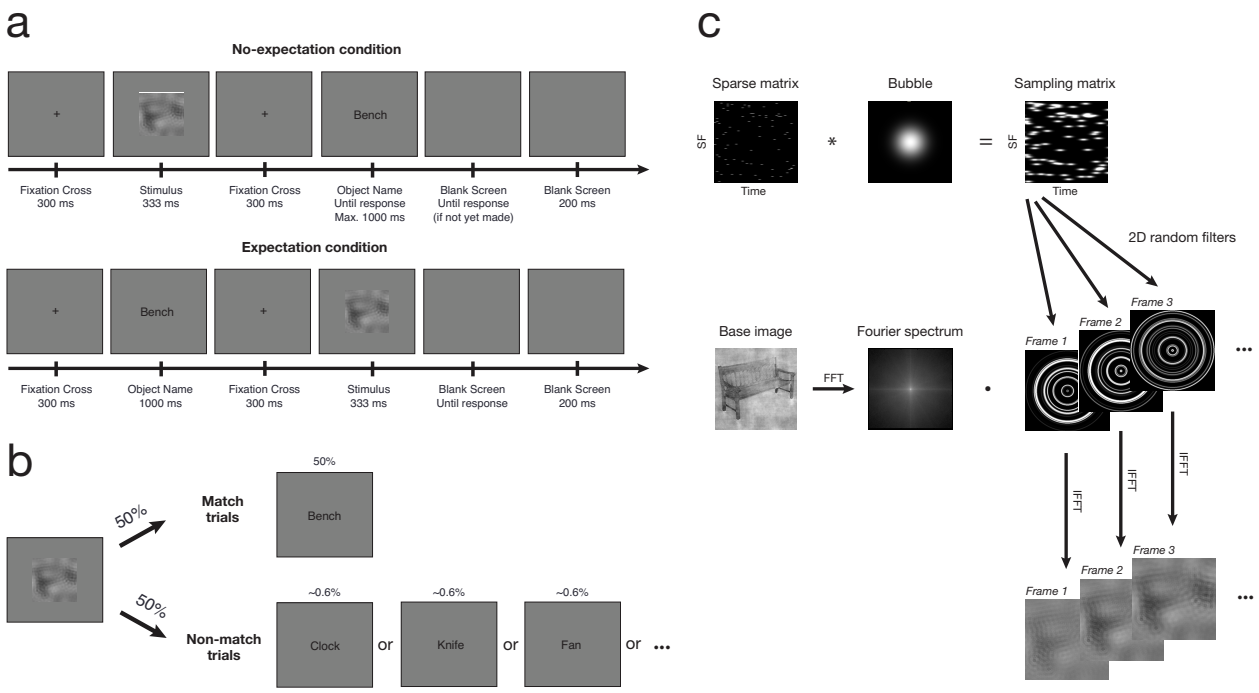


Figure 2.1. Experimental paradigm and sampling method. **A)** On each trial, a fixation cross appeared on a mid-gray background for 300 ms, followed by either an object (no-expectation condition) for 333 ms or an object name (expectation condition) for 1000 ms. Then, another fixation cross appeared for 300 ms, followed by either an object name or an object, depending on what had been presented before. Each trial ended with a blank mid-gray screen. Stimuli, names and fixation crosses have been enlarged for display purposes; names have been translated from French. **B)** Objects and object names had a 50% probability of matching. Since there was a total of 80 objects, each other object had a probability of appearance of about 0.6%. **C)** On each trial, we randomly generated a matrix of dimensions 256×40 (representing respectively SFs and frames) in which most elements were zeros and a few were ones. We then convolved this sparse matrix with a 2D Gaussian kernel (a "bubble"). This resulted in the trial's sampling matrix, shown here as a plane with a number of randomly located bubbles. Every column of this sampling matrix was then rotated around its origin to create isotropic 2D random filters. Finally, these 2D random filters were dot-multiplied by the base image's spectrum and inverse fast Fourier transformed to create a filtered version of the image for every video frame.

2.4.1 Object expectations reduce response times and the amount of information needed for accurate recognition

The mean number of bubbles required to maintain a 75% performance was 45.20 (σ = 15.30) in the no-expectation condition and 33.71 (σ = 10.47) in the expectation condition: the no-expectation condition required significantly more bubbles to attain the same performance ($t(58) = 8.28$, $p = 2.08 \times 10^{-11}$, $d_z = 1.08$). The mean response time was 729 ms (σ = 141 ms) in the no-expectation condition and 598 ms (σ = 183 ms) in the expectation

condition: response times were significantly smaller in the expectation condition ($t(58) = 11.27$, $p = 3.08 \times 10^{-16}$, $d_z = 1.47$).

2.4.2 Coarse information is used earlier when a specific object is expected

To uncover which SFs presented at which moments led to correct responses in each condition, we reverse correlated accuracies with spatial frequencies presented at each moment (see Methods). Mean results are displayed in Figure 2.2: higher z-scores indicate a stronger correlation between the presentation of a given SF on a given time frame and recognition accuracy. In both conditions, a large band of relatively low spatial frequencies (approx. 1–30 cycles per image, cpi) was significant throughout the stimulus duration; however, the peak use of LSFs occurred at distinct latencies when comparing both conditions (Figure 2.2). Specifically, a large significant cluster peaked at 10 cpi and 142 ms ($Z = 11.87$, $p_{FWER} = 1.58 \times 10^{-29}$, $d_z = 1.50$) and at 11 cpi and 232 ms ($Z = 12.04$, $p_{FWER} = 1.95 \times 10^{-30}$, $d_z = 1.44$) in the no-expectation condition. In the expectation condition, a cluster peaked at 11 cpi and 100 ms ($Z = 14.23$, $p_{FWER} = 7.90 \times 10^{-43}$, $d_z = 1.72$) and at 9 cpi and 217 ms ($Z = 14.39$, $p_{FWER} = 8.36 \times 10^{-44}$, $d_z = 1.67$). Contrasting results from both conditions resulted in a significant difference in the same band, peaking at 12 cpi and 67 ms ($Z = 4.03$, $p_{FWER} = .007$, $d_z = .48$): these early low SFs led to more accurate responses in the expectation condition than in the no-expectation condition (Figure 2.2).

To confirm that these low SFs are used by individual subjects earlier in the expectation condition, we conducted an additional latency analysis. We first made the range of values in the classification images (i.e. the maps of regression coefficients) of each subject in each condition vary between 0 and 1 to avoid that a difference in signal-to-noise ratio between our conditions confounds our analysis. We then verified at which latency spatial frequencies under 30 cpi were first above a threshold of .75 in each transformed classification image. A paired t-test revealed that latencies were significantly smaller in the expectation condition compared to the no-expectation condition (38 ms vs 63 ms; $t(58) = 2.77$, $p = .007$, $d_z = .36$).

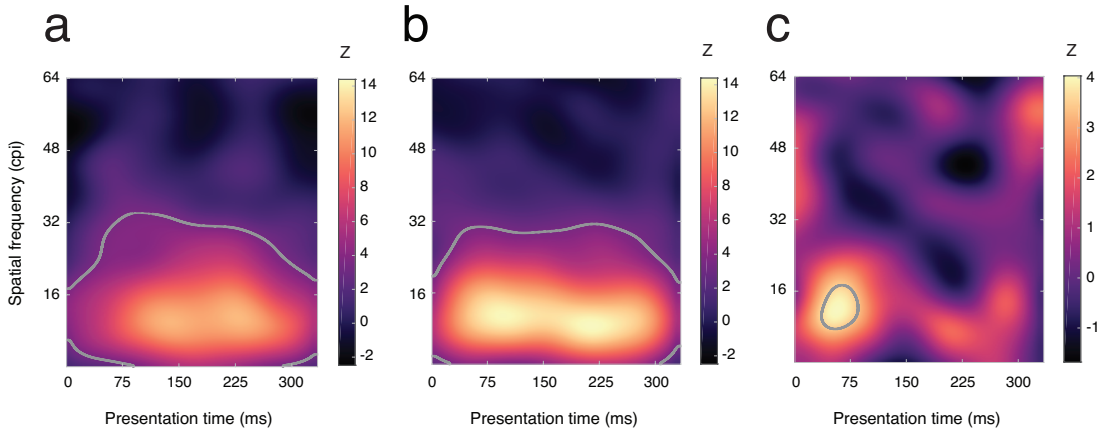


Figure 2.2. Use of SFs across time for each condition and difference between the conditions **A)** No-expectation condition. **B)** Expectation condition. **C)** Expectation – No-expectation. Gray outlines indicate statistical significance at the pixel level ($p < .05$, two-tailed, FWER-corrected). Note that since the width of all images subtended 6 degrees of visual angle, cycles per image can be converted in cycles per degree by dividing by 6.

2.4.3 Expectations modulate information use in an object-specific way

We then investigated whether different expectations for specific objects were modulating the use of SF information differently. To do so, we performed one reverse correlation analysis for match trials of each object in each condition and we computed the variance across the resulting object classification images for each condition (see Methods). There was no significant variance in the no-expectation condition (Figure 2.3a). This was predicted, given that subjects did not know which object would appear in advance and that the amplitude spectrums were equated across objects. In the expectation condition, there was significant across-object variance in low SFs around 9 cpi and 50 ms ($Z = 3.82$, $p_{FWER} = .007$) and around 7.5 cpi and 167 ms ($Z = 6.44$, $p_{FWER} = 1.63 \times 10^{-8}$; Figure 2.3b). In fact, there was significantly more across-objects variance in low SFs around 8 cpi and 175 ms in the expectation condition than in the no-expectation condition ($Z = 4.96$, $p_{FWER} = .0001$; Figure 2.3c). Thus, different object expectations modulate the use of these SFs at that moment differently (see Figure 2.3d and object examples in Figure 2.3e).

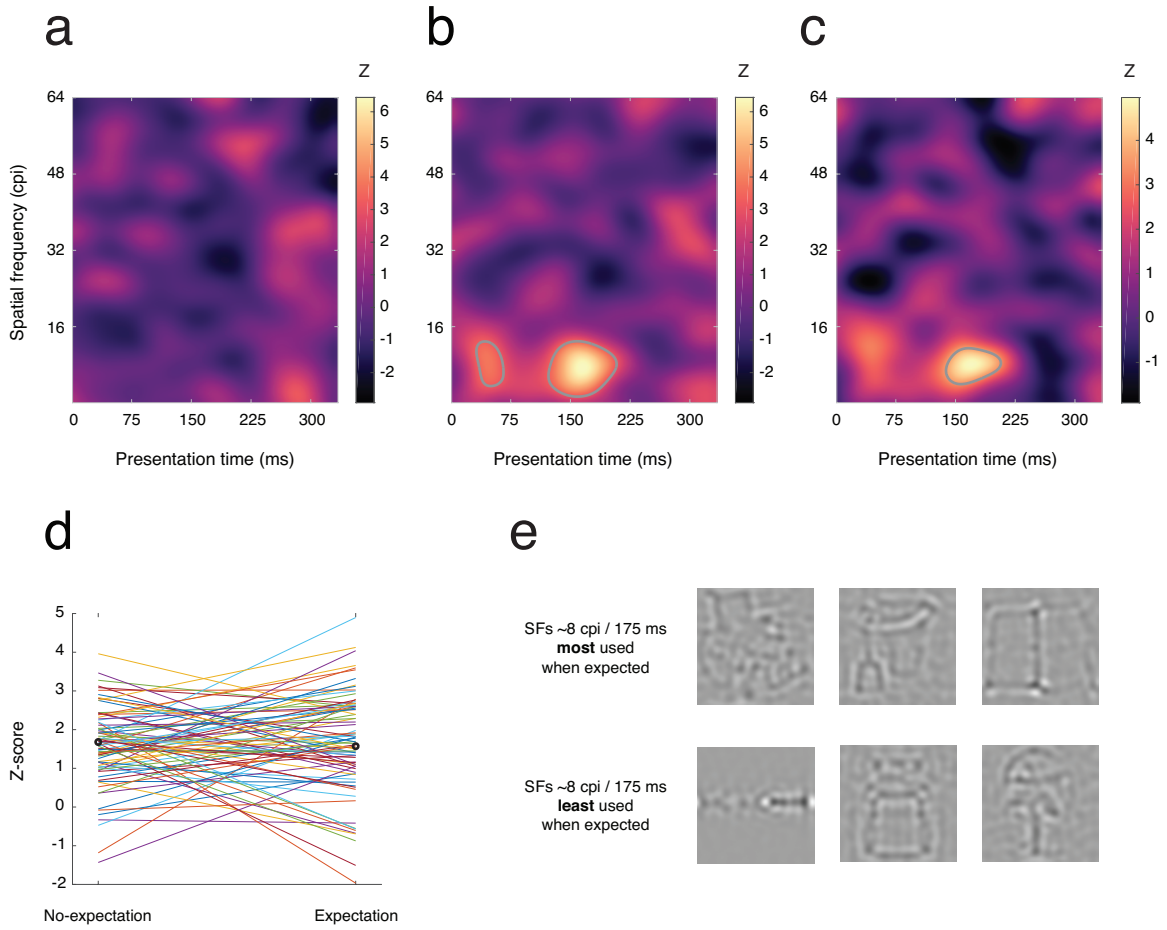


Figure 2.3. Variance across objects in the use of SFs across time for **A)** the No-expectation condition, **B)** the Expectation condition, and **C)** the difference between the conditions (Expectation – No-expectation). Gray outlines indicate statistical significance at the pixel level ($p < .05$, one-tailed for each condition and two-tailed for the difference, FWER-corrected). Note that since the width of all images subtended 6 degrees of visual angle, cycles per image can be converted in cycles per degree by dividing by 6. **D)** Illustration of the use, for each object and condition, of the 8 cpi SFs around 175 ms (the peak activation in C). Black circles indicate the mean for each condition. **E)** Illustration of some of the objects for which the 8 cpi SFs are most and least used around 175 ms, when filtered in the SF band significantly more used in the Expectation condition than in the No-expectation condition (SFs around 8 cpi). Upper row: wheelchair, beach bucket and shovel, book. Lower row: knife, gas pump, fan.

2.4.4 Modulation of information use is temporally nonuniform

So far, we've reported an overall increase in the use of SFs centered on 12 cpi at the start of information acquisition and a variable modulation of the use of SFs centered on 8 cpi mid-way during information acquisition. Next, we investigated whether these two modulation effects were indeed occurring on significantly different SFs and at significantly different time points. A jackknife resampling method (see Methods) confirmed that the

peaks happened at significantly different moments ($Z = 6.37$, $p = 1.85 \times 10^{-10}$), but not that they occurred on significantly different SFs ($Z = 1.47$, $p = .14$).

2.4.5 Late use of detailed information is correlated with recognition ability when no specific object is expected

There is a large variability in object recognition ability across individuals (e.g., 24): skilled observers are likely to have distinct object representations and to exploit expectations more efficiently. We examined whether the use of any specific information was correlated with general object recognition ability in each condition. We computed a sum of the z-scored individual classification images, weighted by the performance indices of the corresponding subjects (calculated as the across-conditions mean numbers of bubbles used to attain a correct response rate of 75% times negative one, z-scored across subjects). In the no-expectation condition, object recognition ability was positively correlated with the use of SFs around 35 cpi in the last stimulus frame ($Z = 3.49$, $p_{FWER} = .044$, $r^2 = .22$; Figure 2.4), suggesting richer representations in skilled observers. In the expectation condition, the use of low SFs around 11.5 cpi and 258 ms was correlated with better general recognition ability ($Z = 4.63$, $p_{FWER} = .0005$, $r^2 = .40$; Figure 2.4). When computing the difference between the conditions, SFs around 35 cpi in the last frame were significant ($Z = -3.54$, $p_{FWER} = .037$; Figure 2.4). Thus, skilled observers appeared to rely on high SFs less in the expectation condition than in the no-expectation condition, suggesting that expectations were used to fill in high SF information.

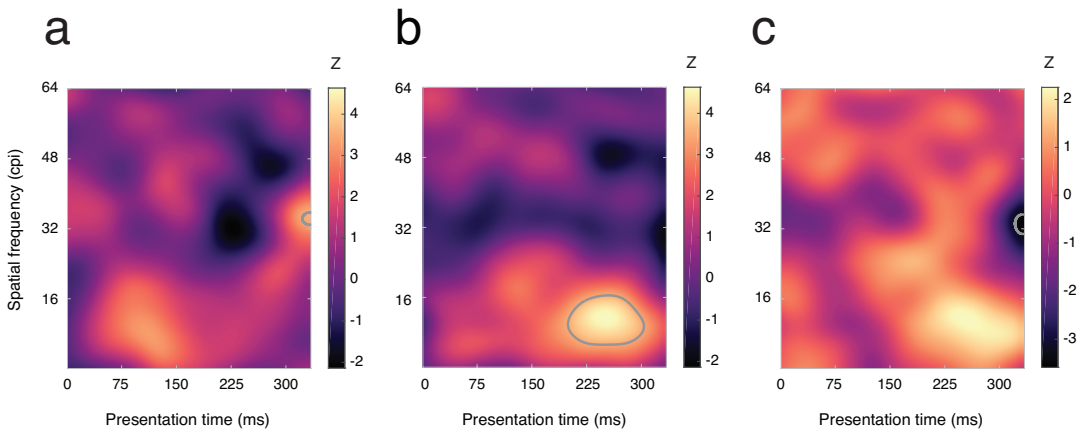


Figure 2.4. Correlation between the use of SFs across time and general object recognition ability for each condition and difference between the conditions **A)** No-expectation condition. **B)** Expectation condition. **C)** Expectation – No-expectation. Gray outlines indicate statistical significance at the pixel level ($p < .05$, two-tailed, FWER-corrected). Note that since the width of all images subtended 6 degrees of visual angle, cycles per image can be converted in cycles per degree by dividing by 6.

2.5 Discussion

In this study, we investigated how expectations of specific objects modulate the use of information for object recognition. We used a large set of complex everyday objects to investigate how the use of different features are affected. Specifically, we examined how the use of SFs across time was affected by expectations. When no specific object was expected, a large band of relatively low SFs (up to about 30 cpi, or 5 cycles per degree, cpd) was used, and their use did not vary across objects. Observers seem to apply a generic recognition strategy in this case, and they use a large portion of the feature space. Such low SFs have been associated with object recognition in prior studies (e.g., Caplette et al., 2016, 2017; Harmon & Julesz, 1973; Hughes et al., 1996). When a specific object was expected, the early use of SFs around 12 cpi (or 2 cpd) was increased on average, likely because subjects were now looking for specific diagnostic features. Several other studies observed heightened sensitivity when stimuli were expected (Cheadle et al., 2015; Stein & Peelen, 2015; Wyart et al., 2012). Expectations further increased the across-objects variability in the later use of similar SFs around 8 cpi (about 1.33 cpd), presumably because the specific object features that were sought at this moment varied in SFs.

We also observed that the late use of higher SFs (around 35 cpi or 5.83 cpd) correlated with object recognition ability when no specific object was expected. A late use of SFs around 35 cpi was also observed in two prior studies using the same method (Caplette et al., 2016, 2017). In these cases, however, it was visible even in the group average; why it is not the case here may be due to the smaller number of trials per participant. When there were specific expectations, the correlation between ability and use of high SFs disappeared. This suggests that expectations can replace the information about the object's identity provided by high SFs in the absence of an expectation. We did not observe, however, a reduction in the use of low SFs, possibly because they are too informative during recognition, especially compared to higher SFs. Thus, expectations do not replace low SF information: this may occur either because low SFs are not used to create hypotheses during recognition (Bar, 2003) or because a prior expectation is not equivalent to a hypothesis created during recognition.

Notably, all three effects occurred at distinct time points. Early in the fixation, object expectations uniformly modulated features prominently represented in low SFs; later, they affected features whose low SF content was highly variable across objects. Later still – in the last stimulus frame –, they reduced the use of high SF features at least in the subjects most proficient at object recognition. This temporal nonuniformity of expectation effects may result from an underlying temporal evolution of information use. If, at any given moment, we use specific features, these are the features whose processing will be modulated by expectations. Such temporally variable information sampling has been observed in several previous studies (e.g., Caplette et al., 2016; Dupuis-Roy, Faghel-Soubeyrand & Gosselin, 2019; Schyns & Oliva, 1994; Vinette, Gosselin & Schyns, 2004). Conversely, our results also illustrate that not all features are affected by expectations in the same way. In addition to varying temporal dynamics, we also show that the use of high SFs at least by the best subjects is *reduced* when there are object expectations.

Few other studies used reverse correlation in the context of expectations, and all used orientation discrimination tasks in which subjects had to choose between two alternatives. In one of these studies, the authors report a shift of the observers' templates toward a more optimal position on the orientation continuum, so that subjects are overall more sensitive when they expect the presented orientation (Cheadle et al., 2015). These

results are compatible with ours, as we also observe an increase in sensitivity when subjects have expectations. In another study, subjects were shown a series of Gabor patches randomly varying in orientation and they had to indicate whether the average orientation was clockwise or counterclockwise (Bang & Rahnev, 2017). Stimuli were either preceded or followed by a cue indicating the most likely orientation category, in a similar fashion to our study. The authors found that pre and post cues produced similar effects. Although this can seem to contradict our results at first, it is difficult to link these findings to object recognition processes occurring in everyday life because of the use of artificial task and stimuli. Furthermore, while expectations in previous studies were directly about features, in our study, they were about complex everyday objects, which vary along multiple feature dimensions. Such predictions are likely to occur more often in a natural environment where most temporal contingencies are high-level: objects are associated with specific objects, contexts and abstract concepts while low-level features predict other low-level features less reliably. In addition, 'mnemonic' expectations based on associations learned a long time ago might rely on different neuronal mechanisms (Caplette, Gosselin, Mermillod & Wicker, submitted; Hindy, Ng & Turk-Browne, 2016).

We have shown how expectations influence the recognition of complex everyday objects through time. In summary, when subjects have no expectations, they sample coarse information throughout recognition – the best subjects also seem to use detailed information at the very end of the fixation; when subjects expect specific objects, they sample object-specific coarse features, and then object-specific features that are not necessarily coarse. These results reveal for the first time the mechanisms underlying the effects of real-world object expectations on their recognition.

2.6 Methods

2.6.1 Participants

Fifty-nine right-handed neurotypical adult participants (29 males; mean age = 22.07; SD = 2.87) were recruited on the campus of the University of Montreal. Subjects had normal or corrected to normal vision and did not suffer from any visual or reading disability. The

study was approved by the ethics board of the University of Montreal's Faculty of Arts and Sciences. Written consent from all participants was obtained after the procedure had been fully explained, and a monetary compensation was provided upon completion of the experiment.

2.6.2 Materials

The experimental programs ran on Mac Pro (Apple Inc.) computers in the Matlab (Mathworks Inc.) environment, using functions from the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard & Pelli, 2007; Pelli, 1997). All stimuli were presented on Asus VG278H monitors (1920 × 1080 pixels at 120 Hz), calibrated to allow a linear manipulation of luminance. Luminance ranged from 1.6 cd/m^2 to 159 cd/m^2.

2.6.3 Stimuli

Eighty grayscale images of man-made objects were selected from the database used in (Shenhav, Barrett & Bar, 2013) and from Internet searches (Caplette et al., 2016, 2017). Images were 256 × 256 pixels and median object width was 220 pixels. The objects were cropped manually and pasted on a homogenous mid-gray background. The Fourier amplitude spectrum of each image was set to the mean Fourier amplitude spectrum across images and the mean luminance of each image was set to the global mean luminance across images using the SHINE toolbox (Willenbockel et al., 2010). Resulting images had a root mean square (RMS) contrast of 0.20.

On each trial, participants were shown a short video (333 ms) consisting of an object image with random SFs gradually revealed at random time points (e.g., Video S1; Video S2); that is, on each video frame, there would typically be several SFs shown among all possible SFs, and these would change from frame to frame. To create these dynamic stimuli, we first randomly generated, on each trial, a matrix of dimensions 256 x 40 (representing respectively SFs from 0.5 to 128 cpi, and frames, each lasting 8.33 ms; the matrix was in fact larger because of additional padding) in which most elements were zeros and a few were ones. The probability of an element being one was adjusted on a trial-by-

trial basis to maintain performance at 75% correct. We then convolved this *sparse matrix* with a 2D Gaussian kernel (a "bubble"; σ_{SF} = 1.25 cpi; σ_{time} = 12.5 ms). This resulted in the trial's *sampling matrix*: a SF × time plane with randomly located bubbles. Every column of this sampling matrix was then rotated around its origin to create isotropic *2D random filters*. Finally, these 2D random filters were dot-multiplied by the base image's Fourier amplitude spectrum and inverse fast Fourier transformed to create a filtered version of the image for every video frame (Figure 2.1c; Caplette et al., 2016, 2017). To ensure accurate luminance display, we applied noisy-bit dithering to the final stimuli (Allard & Faubert, 2008).

2.6.4 Procedure

Participants sat in front of a computer monitor, in a dim-lighted room, over the course of two days. They completed two 500-trial blocks in each experimental session. In half of the trials, an object name was shown before the stimulus (*expectation* condition); in the other half, the stimulus was shown before the object name (*no-expectation* condition). These two conditions were alternated in mini-blocks of 50 trials (the first condition was counterbalanced across participants) and a short pause occurred between mini-blocks. Specifically, each trial was comprised of the following consecutive events on a mid-gray background: a fixation cross (300 ms), a basic-level object name (1000 ms) or the video stimulus (333 ms), a fixation cross (300 ms), the video stimulus (333 ms) or a basic-level object name (until response, max. 1000 ms), and a blank screen (until response if not yet made + 200 ms). When shown after the stimulus, the object name remained on screen either until a response was provided or for a maximum of 1 s, in which case it was replaced by a blank screen until a response was provided; when shown after the name, the video stimulus was immediately replaced by a blank screen until a response was provided (this is in addition to a final 200 ms blank screen after the response; Figure 1a). The number of bubbles was adjusted on a trial-by-trial basis using a gradient descent algorithm within each condition to maintain performance at 75% correct (actual mean performance was 75.90%). Subjects were asked to indicate whether the name matched the object as accurately and as rapidly as possible. The object name and the object matched on 50% of

trials; on the trials in which they didn't match, the name was randomly chosen among the names of all other objects. During the task, chin rests were used to maintain viewing distance at 76 cm; images subtended 6 x 6 degrees of visual angle.

2.6.5 Data Analysis

A few experimental sessions were excluded prior to the analysis: one because of technical problems, three because the mean numbers of bubbles were over three standard deviations above the mean, and two because the mean response times were over three standard deviations above the mean.

Accuracies and response times were z-scored within each object (to minimize variability due to differences in object recognizability), block (to minimize variability due to task learning) and subject (to minimize variability due to interindividual differences in performance), separately for each condition. Trials associated with z-scored response times or accuracies over 3 or below –3 were discarded from the regressions (1.88% of trials). Sparse matrices were also z-scored within each trial. Only spatial frequencies up to 64 cycles per image (cpi) were analyzed, since higher SFs typically do not contribute to accurate object recognition (Caplette et al., 2016, Caplette, West, Gomot, Gosselin & Wicker, 2014; Gold, Bennett & Sekuler, 1999).

To uncover which SFs in which time frames led to accurate object recognition in each condition, we performed sums of z-scored sparse matrices weighted by z-scored accuracies, separately for each subject and each condition (since matrices are random, this is equivalent to a least-square multiple linear regression). The resulting matrices were then summed across subjects and convolved with a Gaussian kernel (σ_{SF} = 6.5 cpi; σ_{time} = 46 ms); henceforth, we will refer to these matrices as classification images. The same procedure was repeated with 500 bootstrapped samples, which were then used to transform the summed regression coefficients into z-scores. Finally, statistical significance of the z-scored classification images was assessed with a two-tailed Pixel Test (Chauvin, Worsley, Schyns & Arguin) which controls the Family-Wise Error Rate (FWER) while taking correlation in the data into account.

To uncover whether different object expectations had a different effect on the use of SFs, we conducted a variance analysis. Specifically, we concatenated the z-scored accuracies and sparse matrices of all subjects together and performed one weighted sum for each object name in each condition. Since we were interested in the facilitation of the recognition of an object by the expectation that this same object would appear, we considered only *match* trials in which the object name corresponded to the object shown. Resulting classification images were convolved with a Gaussian kernel (σ_{SF} = 6.5 cpi; σ_{time} = 46 ms) and variance across objects was computed for each spatial frequency in each time frame for each condition. The same procedure was repeated with 500 bootstrapped samples, which were then used to transform the regression coefficients into z-scores. Finally, statistical significance of the classification images was assessed with a Pixel Test (Chauvin et al., 2005). Since variance can only be positive, statistical tests were one-tailed, except for the difference between the conditions. For display purposes (Figure 3b), object classification images were also z-scored with the help of the bootstrapped samples.

To investigate differences in peak locations, we performed jackknife resampling analyses. Specifically, we performed bicubic interpolations of the subject or object classification images (depending on the concerned analysis) to get more fine-grained estimates, and repeated the analyses while taking one subject or object out at each iteration, estimating the locations of the peaks each time. This allowed us to get an estimate of the variance (or uncertainty) around the peak location (θ): $V_\theta = \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \bar{\theta})^2$, where $\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i$. Then, we compared the peak locations (θ and β) and computed a Z statistic: $Z = \frac{\bar{\theta}-\bar{\beta}}{\sqrt{(V_\theta + V_\beta)}}$. Finally, we used a normal distribution to compute a p-value and assess the statistical significance of the result.

2.7 References

Allard, R., & Faubert, J. (2008). The noisy-bit method for digital displays: Converting a 256 luminance resolution into a continuous resolution. *Behavior Research Methods*, *40*(3), 735–743.

Bang, J. W., & Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Scientific Reports*, *7*(1):17072.

Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *15*, 600–609.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research* Reviews, *36*, 96–107.

Bullier, J., & Nowak, L. G. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology*, *5*(4), 497–503.

Caplette, L., Gosselin, F., Mermillod, B., & Wicker, B. (submitted). Real-world expectations and their affective value modulate object processing.

Caplette, L., Wicker, B., & Gosselin, F. (2016). Atypical Time Course of Object Recognition in Autism Spectrum Disorder. *Scientific Reports*, *6*:35494.

Caplette, L., Wicker, B., Gosselin, F., & West, G. (2017a). Hand position alters vision by modulating the time course of spatial frequency use. *Journal of Experimental Psychology: General*, *146*(7), 917–923.

Caplette, L., West, G., Gomot, M., Gosselin, F., & Wicker, B. (2014). Affective and contextual values modulate spatial frequency use in object recognition. *Frontiers in Psychology*, *5*:512.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*(9), 659–667.

Cheadle, S., Egner, T., Wyart, V., Wu, C., & Summerfield, C. (2015). Feature expectation heightens visual sensitivity during fine orientation discrimination. *Journal of Vision*, *15*(14):14.

Denison, R. N., Piazza, E. A., & Silver, M. A. (2011) Predictive context influences perceptual selection during binocular rivalry. *Frontiers in Human Neuroscience*, *5*:166.

Denison, R. N., Sheynin, J., & Silver, M. A. (2016). Perceptual suppression of predicted natural images. *Journal of Vision*, *16*(13), 6–15.

Dupuis-Roy, N., Faghel-Soubeyrand, S., & Gosselin, F. (2019). Time course of the use of chromatic and achromatic facial information for sex categorization. *Vision Research*, *157*, 36–43.

Esterman, M., & Yantis, S. (2010). Perceptual Expectation Evokes Category-Selective Cortical Activity. *Cerebral Cortex*, *20*(5), 1245–1253.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, *39*(21), 3537–3560.

Harmon, L. D., & Julesz, B. (1973). Masking in Visual Recognition: Effects of Two-Dimensional Filtered Noise. *Science*, *180*, 1194–1197.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665–667.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, *8*(3), 197–230.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, ECVP Abstract Supplement.

Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior Expectations Bias Sensory Representations in Visual Cortex. *Journal of Neuroscience*, *33*(41), 16275–16284.

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554.

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270.

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*, *114*(39), 10473–10478.

Kok, P., Rait, L. I., & Turk-Browne, N. B. (submitted). Content-based dissociation of hippocampal involvement prediction.

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1645–1650.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8):13.

Rajalingham, R., Schmidt, K., & DiCarlo, J. J. (2015). Comparison of Object Recognition Behavior in Human and Monkey. *Journal of Neuroscience*, *35*, 12127–12136.

Schyns, P. G., & Oliva, A. (1994). From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition. *Psychological Science*, *5*(4), 195–200.

Shenhav, A., Barrett, L., & Bar, M. (2013). Affective value and associative processing share a cortical substrate. *Cognitive, Affective, and Behavioral Neuroscience*, *13*, 46–59.

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, *144*(6), 1089–1104.

Sterzer, P., Frith, C., & Petrovic, P. (2008). Believing is seeing: expectations alter visual awareness. *Current Biology*, *18*(16), R697–R698.

Vinette, C., Gosselin, F., & Schyns, P. (2004). Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. *Cognitive Science*, *28*(2), 289–301.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684.

Wyart, V., Nobre, A. C., & Summerfield, C. (2012) Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3593–3598.

2.8 Acknowledgements

The study was supported by a Discovery Grant (436140-2013) from the Natural Sciences and Engineering Research Council of Canada (NSERC) to G. W., by a Discovery Grant (04777-2014) from NSERC to F. G. and by an Alexander-Graham-Bell Doctoral Scholarship from NSERC to L. C.

Chapitre 3

Article 2

Disentangling presentation and processing times in the brain

Laurent Caplette[1*], Robin A. A. Ince[2], Karim Jerbi[1] & Frédéric Gosselin[1]

[1]Department of Psychology, Université de Montréal, Montréal, Qc, Canada

[2]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, United Kingdom

Corresponding author: Laurent Caplette

Email: laurent.caplette@umontreal.ca

Address: Department of Psychology, University of Montreal, C.P. 6128 succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada

3.1 Abstract

Visual object recognition seems to occur almost instantaneously. However, not only does it require hundreds of milliseconds of processing, but our eyes also typically fixate the object for hundreds of milliseconds. Consequently, information reaching our eyes at different moments is processed in the brain together. Moreover, information received at different moments during fixation is likely to be processed differently, notably because different features might be selectively attended at different moments. Here, we introduce a novel reverse correlation paradigm that allows us to uncover with millisecond precision the processing time course of specific information received on the retina at specific moments. Using faces as stimuli, we observed that processing at several electrodes and latencies was different depending on the moment at which information was received. Some of these variations were caused by a disruption occurring 160-200 ms after the face onset, suggesting a role of the N170 ERP component in gating information processing; others hinted at temporal compression and integration mechanisms. Importantly, the observed differences were not explained by simple adaptation or repetition priming, they were modulated by the task, and they were correlated with differences in behavior. These results suggest that top-down routines of information sampling are applied to the continuous visual input, even within a single eye fixation.

3.2 Significance Statement

Typically, studies investigating visual object recognition consider the presentation and perception of an object as a single event. But in fact, it is a series of events: information received on the retina at each moment during fixation is processed, potentially differently, by the brain through time. In this study, we introduce a novel paradigm based on reverse correlation to reveal how information received on the retina at specific moments is processed – we show significant differences depending on when information was received. Our approach allows to visualize a previously unseen dimension to visual processing, and to disentangle information reception and information processing, which are both unravelling through time.

3.3 Introduction

Visual object recognition is a process that seems to occur almost instantaneously. However, this is just an impression: not only does our brain process the object for hundreds of milliseconds, but we will typically fixate it for hundreds of milliseconds too. Because of this, light reflected on an object and reaching our eyes at different moments will typically be processed in the brain at the same moment (but possibly at different processing levels; Figure 3.1). Therefore, brain activity evoked by the perception of an object is a combination of the brain responses to information received on the retina at different moments.

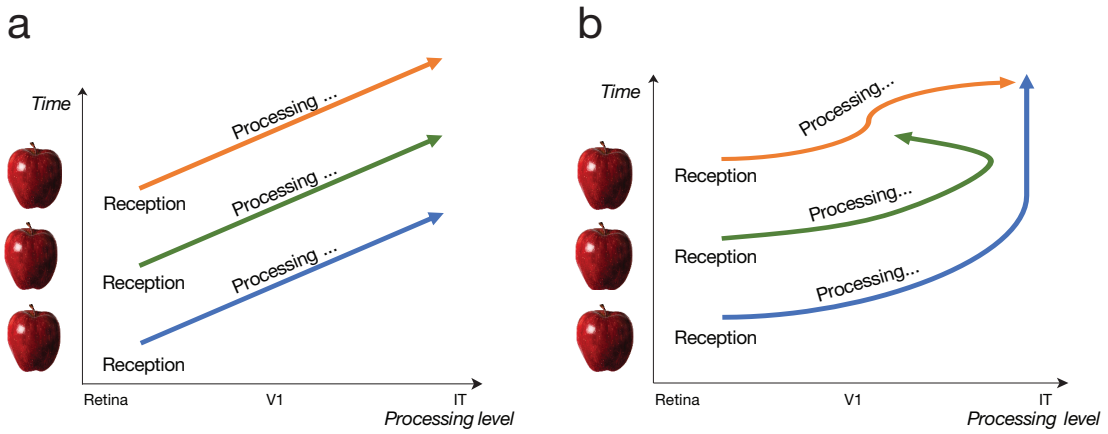


Figure 3.1. At any given point in time (any horizontal imaginary line in the above graphs), information received at different moments during fixation is simultaneously processed in the brain (possibly at different processing levels). **A)** Processing is identical for information received at different moments. **B)** Processing is different for information received at different moments.

We can also expect visual information received at different moments to be processed differently (Figure 3.1b). This is partly because of the limited processing capacity of higher visual areas (Broadbent, 1958; Desimone & Duncan, 1995), which prevents too much information from being processed simultaneously. One strategy that can be applied by the visual system to overcome this limitation is to use visual information received in different time windows to process different features (e.g., different regions of space, colors or spatial frequencies). This is often referred to as top-down attention being guided from one feature to another (Baluch & Itti, 2011; Carrasco, 2011) as a visual routine (Ullman, 1984), or simply as a sampling of different features across time.

The use of the information received at specific moments to process specific features may arise because this is a more efficient strategy for some tasks than using information received at any moment to process any feature (Ullman, 1984). Moreover, specific strategies may be more efficient than others. For example, it may be computationally more efficient to process coarse information before finer noisier features, when recognizing objects or scenes (Marr, 1982; Watt, 1987), and so, high visual areas might process coarse information received early and fine information received late but not fine information received early. It follows that relatively stable strategies may occur in individuals, or even across individuals. Other biases may also result in stable strategies: for example, a tendency to process the most informative features in the information received first (which is probably an evolutionarily sensible strategy), or an attempt to compensate anatomical limitations (e.g., process color from the information received earlier because color is processed more slowly; Bartels & Zeki, 2006; Dupuis-Roy, Faghel-Soubeyrand & Gosselin, 2019). These strategies are likely to depend on the expected input and on the task.

How information received at different moments within a fixation is processed for object recognition is rarely investigated, possibly in part because the distinction between presentation (or reception) time and processing time is not often discussed or appreciated (but see VanRullen, 2011). Still, a few behavioral studies have examined this question, either by randomly revealing image features across time (Blais, Roy, Fiset, Arguin & Gosselin, 2012; Caplette, Wicker & Gosselin, 2016; Caplette, Wicker, Gosselin & West, 2017a; Dupuis-Roy et al., 2019; Vinette, Gosselin & Schyns, 2004) or by adding noise that is randomly varying across time (Nagai, Bennett & Sekuler, 2007; Neri & Levi, 2007), and by correlating the samples with the subject's response. These methods and similar ones (e.g., randomly varying inter-stimulus intervals with high resolution) have been employed several times in the related literature on attention and detection mechanisms (Fiebelkorn, Saalmann & Kastner, 2013; Landau & Fries, 2012; Latour, 1967; Neri & Heeger, 2002; Neri & Levi, 2008; Tse, 2004). Using such methods in object recognition paradigms has led to multiple demonstrations of how observers use the information received at different moments to categorize an object. Interestingly, these strategies often seem stable across individuals. For example, as it was hypothesized, correct responses correlate with high spatial frequency, or fine, information received late, and with low spatial frequency, or

coarse, information received early and late (Caplette et al., 2016, 2017a, 2017b; see also Hughes, Nozawa & Kitterle, 1996; Parker, Lishman & Hughes, 1996). These strategies also seem to be contingent on the task at hand (e.g., Schyns & Oliva, 1999).

While studies have been conducted on the effects of stimulus onset asynchrony (e.g., Bacon-Macé, Macé, Fabre-Thorpe & Thorpe, 2005), duration (e.g., Brisson & Jolicoeur, 2007; Tanskanen, Näsänen, Ojanpää & Hari, 2007), and ordering (e.g., Kauffmann, Chauvin, Pichat & Peyrin, 2015) on brain activity, the processing by the brain of information received at specific *moments* during a fixation has, to our knowledge, never been investigated. This a fundamentally different endeavor: decomposing the processing time course of an object according to the moment at which information is received should inform us about the neural mechanisms underlying the differential sampling and integration of information across time. It should allow us to disentangle the sampling and the processing of visual information, which are both unraveling through time.

In this study, we aimed to perform such a decomposition. To do so, we randomly sampled the features of a face across time while subjects were performing a gender or expression recognition task (Dupuis-Roy et al., 2019; Vinette et al., 2004; see Figure 3.2 and Movies S1-S4) and while their EEG activity was recorded. Faces were chosen as stimuli because they are important social stimuli that human brains are wired by evolutionary pressures to process efficiently; moreover, faces are particularly well suited to a spatial sampling of information as they all are composed of the same spatial features with essentially the same spatial configuration. To ensure that subjects could initiate a potential top-down sampling strategy on time, face stimuli occurred at predictable moments. We then reverse correlated brain activity at all time points to information presented in different time windows. We had three main hypotheses: 1) the processing time course of information received at different moments will be different; 2) this modulation of processing for different information reception moments will itself be modulated by the task; and 3) variations in the processing of information received at different moments will correlate to variations in the use of this information for the task.

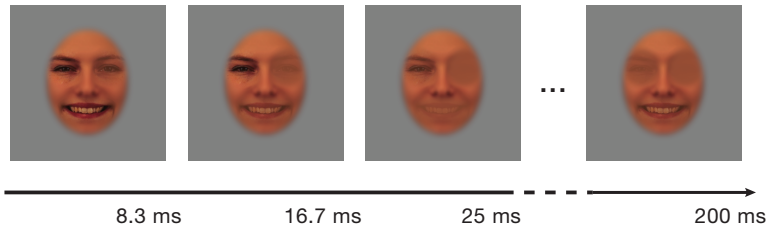


Figure 3.2. Example of a video stimulus used in a random trial. The three face features were smoothly revealed in random frames (1 frame each 8.3 ms) across 200 ms. See movies S1-S4.

3.4 Results

3.4.1 Time course of information use

Mean accuracy was 75.8% (σ = 4.2%) in the gender task and 82.9% (σ = 6.2%) in the expression task. Mean response time was 711 ms (σ = 87 ms) in the gender task and 662 ms (σ = 100 ms) in the expression task.

To identify which face features in which time frames led to accurate responses, we performed for each session a sum of sampling matrices (indicating the visibility of each face feature at each time frame in the stimulus on each trial) weighted by accuracies. Mean results for each task are displayed in Figure 3.3. As we can see, both eyes were used at all except the earliest reception moments (i.e. time points at which information is presented and received on the retina) to correctly identify the gender of the faces, while the mouth was used throughout the presentation to identify the expression of the face. These results replicate previous studies using a spatial sampling of the whole face (Dupuis-Roy et al., 2019; Dupuis-Roy, Fortin, Fiset & Gosselin, 2009; Gosselin & Schyns, 2001; Schyns, Bonnar & Gosselin, 2002).

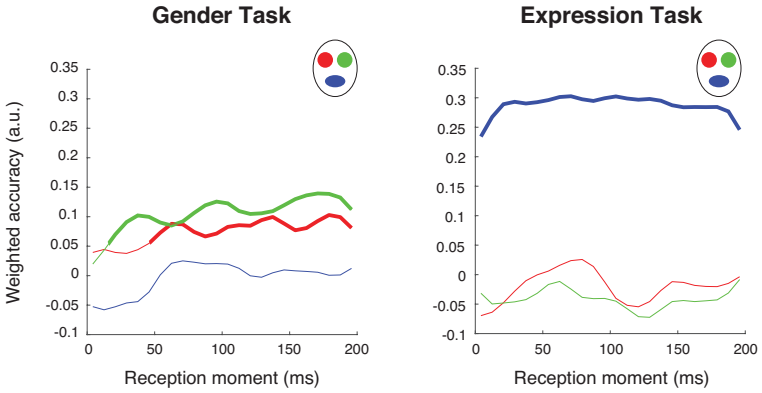


Figure 3.3. Behavioral results indicating, for each task, how each feature presented on each frame correlates with correct responses. Bold segments of line indicate frames that are significant ($p < .05$, one-tailed, FWER-corrected).

3.4.2 Visual Evoked Potentials

To verify if our sampling method elicited, on average, similar ERPs to whole unaltered faces, we computed the average of all trials with sampled and whole faces, for those subjects who performed the task on both kinds of trials. As we can see, ERPs and their associated topographies are very similar (Figure 3.4). We computed the correlation between the two ERPs on peak left and right occipito-temporal sensors (LOT and ROT respectively; see Methods), considering all time points from 80 to 400 ms after stimulus onset, to verify if the patterns of activation irrespective of global amplitude were the same: correlations were very high (LOT: 0.97; ROT: 0.99). This suggests that our sampling method did not greatly alter the average brain response to faces.

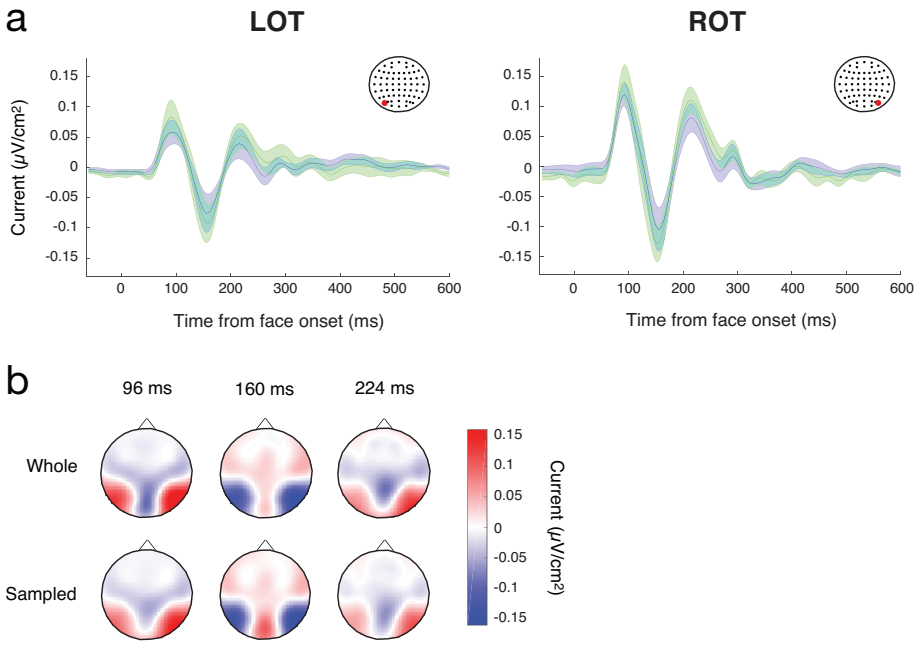


Figure 3.4. A) Mean ERPs for whole (green) and sampled (blue) faces on LOT and ROT. Shaded areas represent standard errors above and below the mean. **B)** Topographies for whole and sampled faces at selected latencies.

3.4.3 Uncovering the processing of information received at different moments

For each session, ridge regressions were performed between sampling matrices of correct trials and EEG amplitude on each time point and electrode (see Methods; Figure S1a). Although analyses were conducted on all electrodes (and appropriate corrections for multiple comparisons were applied), we will mostly focus on results from occipito-temporal sensors. Mean maps of regression coefficients are displayed for representative left and right occipito-temporal sensors (LOT and ROT) on figures 3.5 (gender task) and 3.6 (expression task). These maps show a complete portrait of what is happening during visual recognition: how information received on the retina at different moments throughout fixation is simultaneously processed through time in the brain.

We can immediately see on most maps (especially the ones for the mouth and the contralateral eyes) a clear diagonal trend: as it could be expected, information received *x* ms later is on average processed *x* ms later in the brain. This processing takes the form, in most cases, of a positive activation followed by a negative one and another positive one (analogous to the classic P1, N170 and P3 components). However, there also seem to be

important differences in amplitude across reception moments. To assess whether these differences are statistically significant, we conducted a task × reception moment ANOVA on regression coefficients for each face feature, electrode and EEG latency, after having realigned each row of the previous maps so that the zero point on the *x* axis is the feature onset rather than the face onset (see Methods; Figure S1b).

Significant modulation of processing by the reception moment is visible during almost all the analyzed time window (~50-360 ms; Figure 3.7). Differences are strongest on occipito-temporal sensors, but they are also present on central and frontal sensors, especially at higher latencies (e.g., there is a significant effect of reception moment peaking between 300 and 350 ms on frontal Fpz sensor).

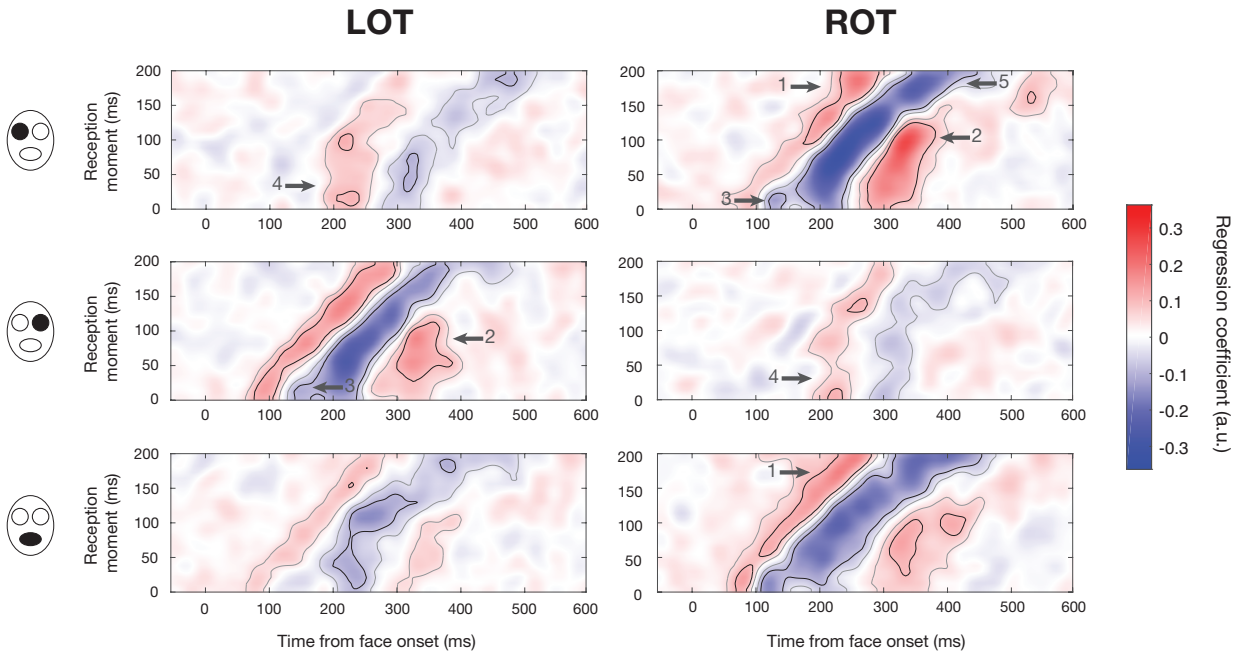


Figure 3.5. Mean maps of regression coefficients for the gender task on LOT and ROT sensors. Gray outlines indicate significance at the cluster level and black outlines indicate significance at the pixel level ($p < .05$, two-tailed, FWER-corrected). Arrows indicate results of interest (see text for details).

On occipito-temporal sensors, variations in the amplitude of the first positive activation across reception moments are leading to significant differences around a latency of 80-100 ms: specifically, this activation is stronger at late reception moments or at all except intermediate reception moments (Arrows #1, Figures 3.5-3.6). The last positive activation peaking at intermediate reception moments is also a source of significant variations around 300 ms (Arrows #2, Figures 3.5-3.6).

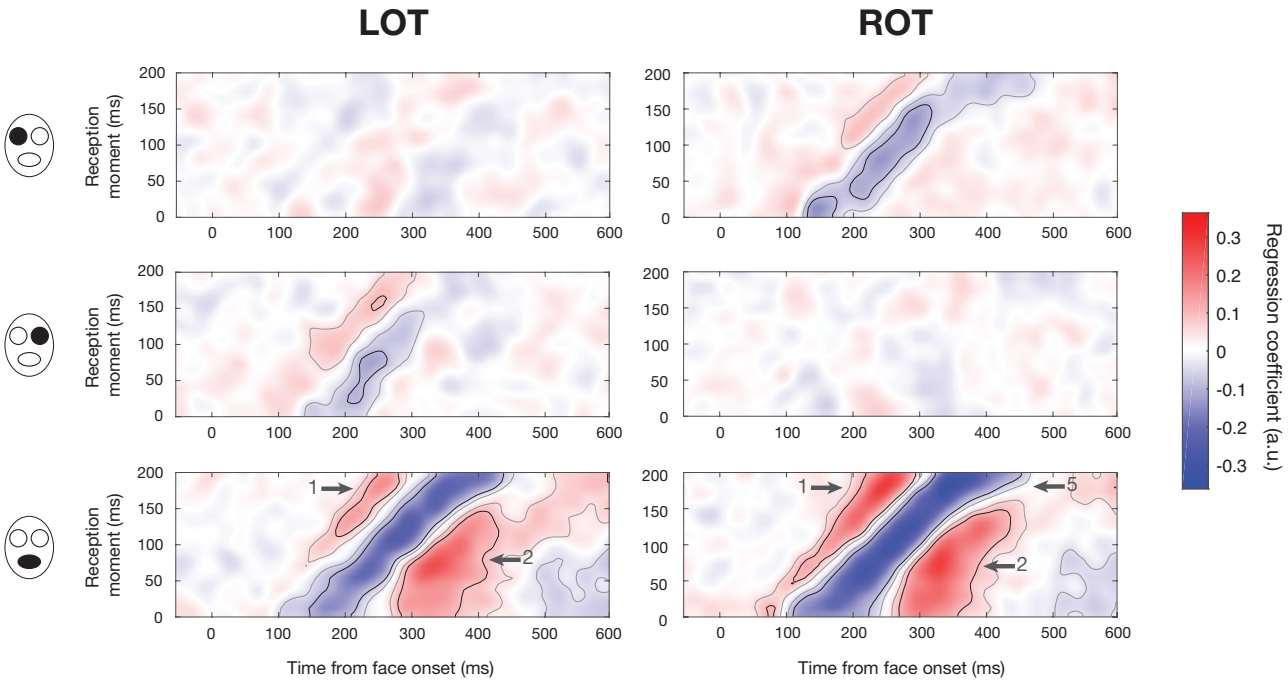


Figure 3.6. Mean results for the expression task on LOT and ROT sensors. Gray outlines indicate significance at the cluster level and black outlines indicate significance at the pixel level ($p < .05$, two-tailed, FWER-corrected). Arrows indicate results of interest (see text for details).

Interestingly, significant differences in amplitude around 150 ms for some feature/task combinations are partly driven by the presence of an apparent additional peak, for the early reception moments (Arrows #3, Figure 3.5). We verified whether these two peaks represented two distinct components with different topographies. To do so, we used the maps of regression coefficients for individual sessions and computed the topographies associated with both peaks (at the same reception moment); we analyzed the 22 subjects who had 2 sessions of data. We thus had four topographies per subject: one for each peak in each session. For each subject, we computed Pearson correlations between topographies associated to the same peak on different days and averaged them: this is the within-peak correlation. Next, we computed Pearson correlations between topographies associated to different peaks on different days and averaged them: this is the between-peaks correlation. We finally performed t-tests between Fisher-transformed within-peak and between-peaks correlation coefficients: the within-peak correlation coefficients were significantly greater (left eye: $t(21) = 3.82$, $p_{Bonf} = .004$; right eye: $t(21) = 3.98$, $p_{Bonf} = .003$). When using the

topographies associated to different peaks on the *same* day to compute the between-peak correlation, we still obtained significantly greater within-peak correlations (left eye: $t(21) = 3.03$, $p_{Bonf} = .03$; right eye: $t(21) = 3.52$, $p_{Bonf} = .008$). In other words, topographies associated with the same peak obtained on different days are more similar than topographies associated to different peaks, even when these are obtained on the same day. Consequently, each peak represents a distinct activation with its own topography and possibly its own neural generators, with the first one being especially sensitive to the onset and stopping being receptive after only about 20 ms.

Other variations on occipito-temporal sensors seem to be driven by increases or decreases in the latency of a component across reception moments. To investigate this, we computed, for each major component, task and feature, the peak latency at each significant reception moment on LOT and ROT (significance at the cluster level; ignoring activations past 500 ms from the face onset). We then fitted a line across these latencies and tested (one-sample t-test) whether the slope of the line was significantly different from one. Here, a slope of one would mean that the feature takes the same time to be processed at all reception moments, whereas a larger slope

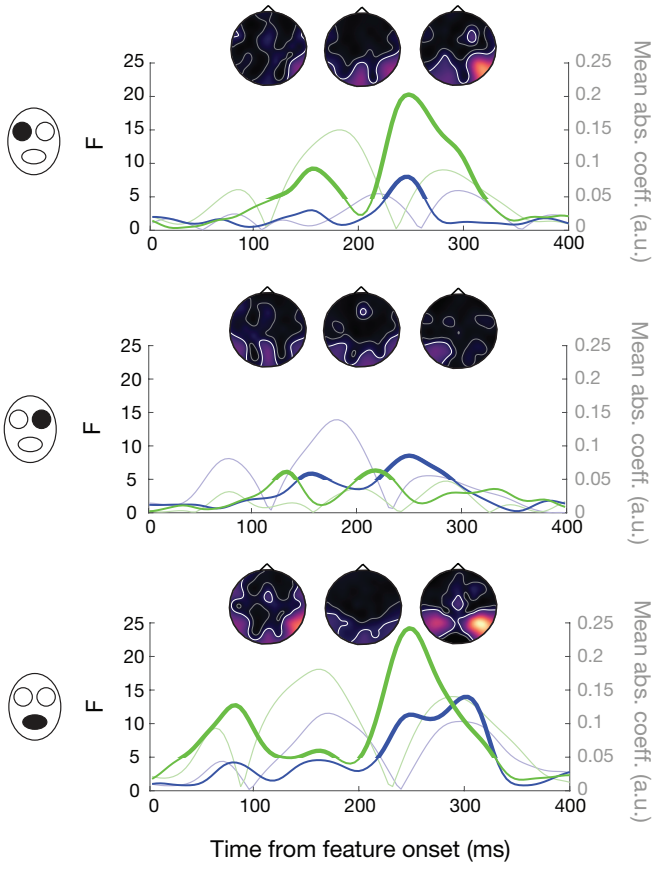


Figure 3.7. Effect of presentation moment on EEG activity, for each face feature. F values are shown for all latencies (from the feature onset) for LOT (blue) and ROT (green) sensors; bold segments indicate time points significant at the pixel level ($p < .05$, FWER-corrected across sensors and time). These time courses are superposed to the mean magnitudes (across reception moments) of regression coefficients for each time point from the feature onset (in smaller point and less saturated color). Topographies depict the temporal progression across the whole scalp: latencies of 100, 150 and 250 ms are shown. White curves indicate areas significant at the pixel level and gray curves indicate areas significant at the cluster level ($p < .05$, one-tailed, FWER-corrected across topography and time).

would mean that the feature takes increasingly longer to be processed with increasing reception moment, and a smaller slope that the feature takes an increasingly shorter time to be processed with increasing reception moment; a slope of zero would mean that features are processed at the same moment irrespectively of when they were received on the retina. In most cases, the latency of the first positive component from the feature onset was approximately constant (i.e. same processing duration for all reception moments) (slopes between .90 and 1.04, $R^2_{adj} > .96$, $df \geq 11$, $t < 2.92$, $p_{Bonf} > .10$) except in the case of the right eye on LOT in the gender task, where it was slightly increasing (slope = 1.08, $R^2_{adj} = .99$, $t(22) = 3.24$, $p_{Bonf} = .049$) and in the case of the eyes on ipsilateral electrodes in the gender task where it was decreasing (slopes < .44, $R^2_{adj} > .27$, $df \geq 17$, $t > 8.84$, $p_{Bonf} < 1.2 \times 10^{-6}$). The small slope for the eyes on ipsilateral electrodes illustrates the striking fact that this component always occurs about 220 ms after the face onset or later; information received the earliest is thus processed at about the same time as information received 50-75 ms later (Arrows #4, Figure 5). Regarding the middle negative component, its slope across reception moments was not different from 1 in most cases (slopes between .60 and 1.44, $R^2_{adj} > .45$, $df \geq 16$, $t < 3.00$, $p_{Bonf} > .08$) except for the left eye on ROT in the gender task and for the mouth on ROT in the expression task (slopes > 1.69, $R^2_{adj} > .78$, $t(22) > 3.68$, $p_{Bonf} < .02$). In both these cases, the slope was significantly larger than one. This is mostly a consequence of an increase in latency in the last reception moments (Arrows #5, Figures 3.5 and 3.6). Finally, in the case of the last positive component, the slope was significantly smaller than 1 for the eyes on the contralateral electrodes in the gender task and for the mouth on LOT in the expression task (slopes between .26 and .66, $R^2_{adj} > .66$, $df \geq 13$, $t > 5.98$, $p_{Bonf} < 2.0 \times 10^{-4}$) and it was approximately constant for the mouth in the gender task and on ROT in the expression task (slopes = .67 and .79, $R^2_{adj} > .66$, $df \geq 11$, $t < 3.45$, $p_{Bonf} > .07$).

3.4.4 Investigating top-down modulations

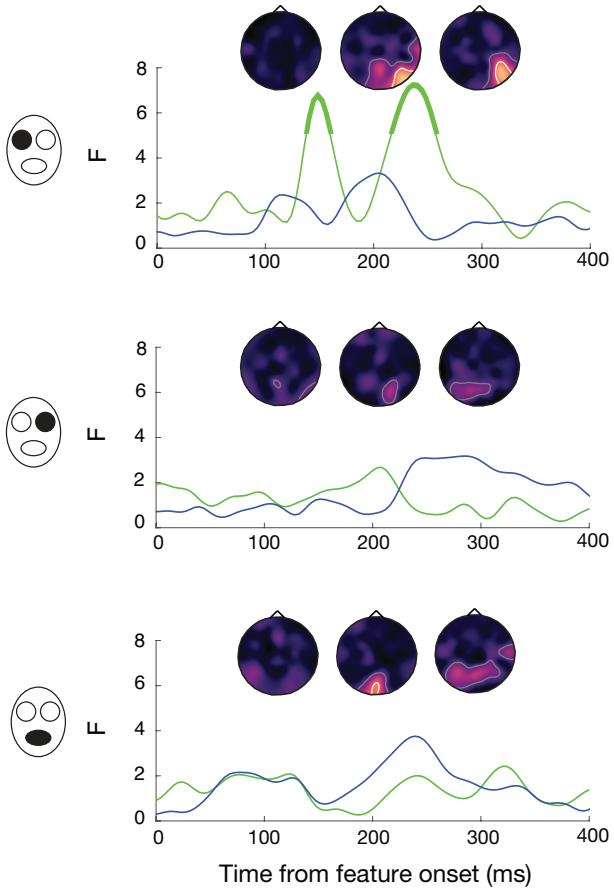


Figure 3.8. Interaction of presentation moment and task on EEG activity, for each face feature. F values are shown for all latencies (from the feature onset) for LOT (blue) and ROT (green) sensors; bold segments indicate time points significant at the pixel level ($p < .05$, FWER-corrected across sensors and time). Topographies depict the temporal progression across the whole scalp: latencies of 100, 150 and 250 ms are shown. White curves indicate areas significant at the pixel level and gray curves indicate areas significant at the cluster level ($p < .05$, one-tailed, FWER-corrected across topography and time).

The differences in processing across reception moments that we uncovered cannot be caused by differences in *what* has been seen before during a trial since sampling was random; however, *how much* was seen could have an influence, since the probability of already having shown information in a trial is greater in the last stimulus frame than in the first one. Thus, the observed differences could be caused in part by bottom-up effects such as adaptation or repetition priming. To investigate this possibility, we repeated the previous regressions only with trials in which just one bubble was revealed: despite a greatly reduced number of trials, results were remarkably similar (Pearson correlation of 0.95 between the maps of regression coefficients; Figures S2 and S3), suggesting that the previously observed effects are not caused by differences in the amount of information perceived beforehand.

This result alone does not completely exclude the possibility of bottom-up effects however. To investigate whether differences in activity across reception moments could be explained at least in part by top-down mechanisms, we verified for each face feature, time point and location, whether there was a significant interaction between reception moment and task, i.e. if the moment at which information is received modulates processing

differently depending on the task. There was a significant interaction at several time points and locations, again mostly on occipito-temporal electrodes but also in more anterior locations. Contrary to what we observed with the main effect of reception moment, there is almost no significant interaction around 100 ms, but the peak effects are similarly around 150 and 250 ms on right occipito-temporal sensors (Figure 3.8). Note that on some more anterior sensors such as CP1, significant interactions peaked after 300 ms.

3.4.5 Relating sampling in the brain and in behavior

We evaluated where and when variations in brain activity across reception moments are related to variations in the behavioral use of information. Since differences in brain activity are likely related to the behavioral use of information in complex nonlinear ways, the mutual information (MI) metric was used. MI was computed across reception moments between coefficients resulting from the accuracy-weighted sums of sampling matrices (behavioral results) and the magnitudes of brain regression coefficients for each subject, face feature, latency from feature onset and electrode. Importantly, computing MI separately for each face feature allowed us to isolate the contribution of *within-feature* variations across reception moments. We observe significant MI mostly on occipito-temporal sensors at early and late latencies, but also in more anterior locations at later latencies (Figure 3.9). Regarding the eyes, significant MI is present early (<130 ms) and

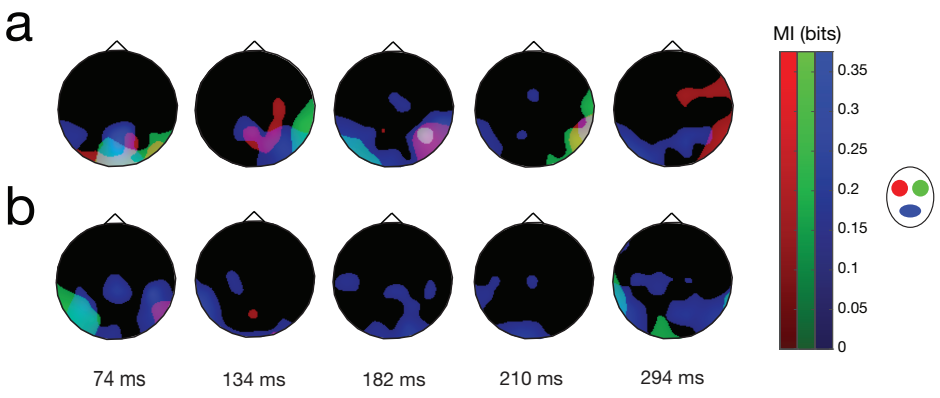


Figure 3.9. Mutual information (MI) between behavioral and brain coefficients (see text for details), for selected latencies. Areas significant at the cluster level are shown ($p < .05$, one-tailed, FWER-corrected across topography and time). Each face feature is represented by a different color channel (red = left eye, green = right eye, blue = mouth) with their combinations representing feature combinations (yellow = both eyes, magenta = left eye and mouth, cyan = right eye and mouth, white = all features); color saturation indicates the magnitude of MI. **A)** Gender task. **B)** Expression task.

late (>250 ms) in both tasks, but it is present at intermediate latencies (~150-250 ms) only in the gender task. Interestingly, significant MI for the mouth is visible throughout the time course, for both tasks. While we did not uncover a significant behavioral use of the mouth in the gender task in our study, other studies have observed it, sometimes only when correlating feature visibility with response times instead of accuracy (Dupuis-Roy et al., 2009; Gosselin & Schyns, 2001; Schyns et al., 2002). These results show that the origin of the variations in the use of information across reception moments can be traced back to variations in occipito-temporal activity at early and late latencies, and to variations in frontal activity at later latencies.

3.5 Discussion

When we fixate an object, light impinges on our retinas in a continuous fashion, implying that our brain processes information that is received at different moments simultaneously, through time and cortical space. This is not typically considered in studies investigating the processing of visual objects, and so the processing uncovered in those studies corresponds to a combination of responses to information received at different moments. In our experiment, we randomly sampled the features of a face across time (Vinette et al., 2004) while brain activity was being measured to decompose this processing and uncover for the first time the brain activity related to information received at specific time points during a single eye fixation.

We first observed that information is processed differently depending on when it is received during fixation. One of the most striking differences is seen in the ipsilateral representation of the eyes on occipito-temporal sensors in the gender task. The lateralized anatomy of the visual system tells us that each eye should be processed by the contralateral hemisphere first (Saenz & Fine, 2010; Van Essen, Newsome & Bixby, 1982): the ipsilateral representation is likely to have been transferred from an early contralateral representation (Ince et al., 2016). Here, the contralateral representation appears to peak at a relatively constant offset of ~175 ms after information is received on the retina, independently of *when* it is received during the stimulus presentation (see the diagonal linear trend of the negative activations in Figure 3.5). However, the ipsilateral representation appears to be

gated: all information received in the first 50 ms of fixation is represented at the same time, around 220 ms from face onset, while information received after 50 ms is represented with a fixed offset of ~120 ms, representation moment increasing linearly with reception moment as for the contralateral representation. Bearing in mind the fact that ipsilateral features must be first processed by the contralateral hemisphere, this suggests that around 220 ms, broadly consistent with the tail end of the classical N170 ERP event (see Figure 3.4), a channel is opened through which features can be transmitted across hemispheres. The N170 has been demonstrated to reflect cross-hemispheric transfer of visual features, with the peak ipsilateral representation of the eyes occurring after the contralateral peak of the N170 event (Ince et al., 2016). The linear relationship between reception moment and representation moment after this gating event suggests that the channel remains open during the remainder of fixation. Despite the same experimental stimuli, this gating phenomenon is only seen in the gender task, suggesting that it is specific to lateralized task-relevant features (the eyes being used almost exclusively for the gender task). In a recent study, the N170 also appeared to filter out task-irrelevant features: while both task-relevant and task-irrelevant features were processed prior to 170 ms, only task-relevant features were processed afterwards (Zhan, Ince, van Rijsbergen & Schyns, 2019). Of note, the cause of this gating cannot be repetition priming because it is also visible in trials where only one feature is revealed once.

Another notable result is the occurrence of two negative peaks instead of one in the contralateral representation of the eyes in the gender task, with the first one sensitive to only a narrow time window after the stimulus onset. Interestingly, these two peaks have significantly distinct topographies, suggesting distinct neural generators. These generators might resemble the generators of the N170 since the activations are similarly peaking around 170 ms after the reception of eye information. Other studies have observed multiple peaks at the expected timing of the N170 (Di Russo et al., 2012; Suzuki & Noguchi, 2013); these are likely corresponding to activity from different generators. In one study, negative peaks around 160 ms have been found to originate from the fusiform gyrus while negative peaks around 180 ms have been localized as originating from the intraparietal sulcus (Di Russo et al., 2012). Interestingly, if we exclude the first peak and only look at the biggest negative cluster, we notice a pattern that is similar to the positive cluster on the ipsilateral

electrodes: all information received in the first ~50 ms is processed at about the same moment (peak around 200 ms) while information received afterwards is processed with a relatively constant (but slightly increasing) offset of 150-170 ms, representation moment increasing with reception moment. It is possible that a gating event occurs here too, preventing processing by the sources of this component to start before ~200 ms after the face onset. This gating occurs at about the same latency as the ipsilateral gating, at the expected timing of the classical N170 ERP component.

Other differences in processing across information reception moments are also visible. For example, the negative activation on ROT has an increased latency for late reception moments for some feature/task combinations (that is, this activation peaks after a longer time interval following the reception of information, if this information is received later). This may be a consequence of the prioritization of information received earlier. The visual system is likely to prioritize information received early since it might be unknown for how long information from that stimulus will reach the retina. Thus, the processing of information received late is likely to be delayed or processed more slowly. The opposite phenomenon was visible for the last positive activation in some cases: its latency was greater at early reception moments. In other words, there was "temporal compression": information received earlier was "maintained" for a longer time and all information was processed at almost the same moment independently of when it was received on the retina. It is expected that information received at different moments is processed simultaneously at some point in the brain if it is to be integrated together by higher level areas. The temporal compression we observe may be a consequence of this process of accumulation and integration of information. This is consistent with other studies reporting a component at similar latencies associated with accumulation of evidence and temporal integration (O'Connell, Docree & Kelly, 2012; Twomey, Murphy, Kelly & O'Connell, 2015).

Although adaptation or priming to previously seen features can be ruled out as a source of these differences because they are also present in trials with only one bubble, a bottom-up cause still might have been possible. For instance, different parts of the visual field may always be processed at specific moments during fixation. To investigate whether there were top-down origins to the effects we observed, we verified whether the task modulated them. We found significant interactions between information reception moment

and task on several sensors at many latencies. In other words, the differences observed in the processing of information received at different moments were not the same depending on the task: consequently, these differences are at least partly top-down in origin. Significant interactions were observed at electrodes and latencies similar to those of the significant effects of reception moment but started slightly later, a result that is expected for top-down modulations. Moreover, significant interactions were occurring in slightly different areas. For example, while the processing of the mouth was globally more modulated by reception moment on right occipital electrodes, the interaction with the task was stronger on central and left occipital electrodes. This suggests that bottom-up mechanisms and top-down sampling are taking place in different loci.

That the brain processes information differently according to when it was received during fixation, that this occurs even when only one such information is revealed in the course of a trial, and that these differences are modulated by the task, all suggest that each time slot is assigned a different "role" in a top-down fashion. This is compatible with the idea of ballistic visual routines: different operations may be applied to the visual input in a sequential fashion, these operations may vary according to the goal of the computation, and the outcome of the first steps does not change the operations applied thereafter. A non-uniform time course of the behavioral use of information in visual recognition has been observed in a few studies (e.g., Dupuis-Roy et al., 2019; Neri & Levi, 2007; Vinette et al., 2004); here, we demonstrate it in the brain for the first time and we show that it is at least partly top-down in origin. Moreover, the variations in processing across reception moments relate to variations in behavior; that is, as it could be expected, how the brain (particularly occipito-temporal areas) processes information received at a specific moment relates to how this information will be used to perform the task.

In summary, we uncovered in this study the neural response to specific information received at specific moments during fixation and we showed that when light is received on the retina matters: processing is modulated by the specific moment at which information is received, even within a single eye fixation. These differences can be quite striking, such as an additional delay of 100 ms for information received at some moments. Importantly, these variations remain even when we account for information perceived beforehand, and they are modulated by the task. Moreover, they correlate to differences in the use of

information for the task. These results suggest that task-dependent visual routines of information sampling are applied top-down to the continuous visual input.

The novel method introduced in this article also seems a promising avenue to shed light on the accumulation and integration of information occurring during object recognition: indeed, it should allow us to visualize the processing, at a given time point and location, of information that was received on the retina at different time points. Future studies using more spatially resolved brain imaging methods such as MEG should investigate how information received at different moments is processed, accumulated, integrated and transferred across brain regions. This method could also be used with intrinsically dynamic stimuli such as dynamic facial expressions or naturalistic movies to investigate how an observer integrates evolving information.

3.6 Methods

3.6.1 Participants

Twenty-four neurotypical adults (mean age = 23.0 years; SD = 2.9) were recruited on the campus of the University of Montreal. Participants did not suffer from any psychiatric or psychological disorder and had no known history of head concussions. The experimental protocol was approved by the ethics board of the Faculty of Arts and Sciences of the University of Montreal and the study was carried in accordance with the approved guidelines. Written informed consent was obtained from all the participants after the procedure had been fully explained, and a monetary compensation was provided upon completion of each experimental session.

3.6.2 Materials

The experimental program ran on a Ciara Discovery computer with Windows 7 in the Matlab environment, using custom scripts and functions from the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard & Pelli, 2007; Pelli, 1997). Stimuli were shown on an Asus VG278H monitor, calibrated to allow a linear manipulation of luminance, with a

resolution of 1920 × 1080 pixels and a 120 Hz refresh rate. Luminance values ranged from 2.47 cd/m^2 to 269 cd/m^2. A chin rest was used to maintain a viewing distance of 76 cm. EEG activity was recorded using an ANT Neuro Waveguard 64-electrode cap with Ag/AgCl electrodes, using a sampling rate of 1024 Hz. Linked mastoids served as initial common reference. Vertical electro-oculogram (vEOG) was bipolarly registered above and below the dominant eye and horizontal electro-oculogram (hEOG) at the outer canthi of both eyes.

3.6.3 Stimuli and sampling

Two hundred and sixty-four color images of faces were selected from the image database *Karolinska Directed Emotional Faces* (Goeleven, De Raedt, Leyman & Verschuere, 2008); only faces facing the camera were chosen. These were composed of 66 different identities (33 women and 33 men) each performing a happy and a neutral expression; two different pictures of each facial expression were used. Faces were aligned on twenty hand-annotated landmarks averaged to six mean coordinates for left and right eyes, left and right eyebrows, nose and mouth, using a Procrustes transformation.

We then created an uninformative face background by taking the mean of all aligned faces and applying a lightly smoothed elliptical mask (horizontal radius = 6 degrees of visual angle) to conceal the background, hair and shoulders. The areas including and surrounding the eyes and eyebrows were then covered by two lightly smoothed approximately circular masks; the area including and surrounding the mouth was covered by a lightly smoothed elliptical mask. The color of these masks was the mean color of the unmasked parts of the average face. The three feature masks were of equal area (within a <1% margin; since feature masks were smoothed, area covered was computed by summing the mask pixel values).

For use in the sampled-face trials, the mean luminance and the contrast of all aligned faces (within the feature areas determined by the feature masks previously discussed) were equalized, separately for each color channel, using the SHINE toolbox[46]. The same procedure was applied but for the whole face (inside the elliptical mask), for use in the whole-face trials.

On each sampled-face trial, the face features of a randomly selected exemplar face were gradually revealed at random moments across a total duration of 200 ms; that is, masked feature areas of the uninformative face background were replaced by the features of an exemplar face (Figure 3.2; Movies S1-S4). A duration of 200 ms was chosen so that no saccade would occur during stimulus presentation on most trials. Specifically, on each trial, a random 3 × 72 sparse matrix composed of zeros and a few ones (the probability of each element being one was constant and was 0.025) was created; each row of 72 elements was then convolved with a 1-D gaussian kernel, or "bubble" (Gosselin & Schyns, 2001; Vinette, Gosselin & Schyns, 2004), with a 1.8 frame (15 ms) standard deviation. Superfluous padding was removed so that the final smoothed matrix was 3 × 24 in size and thresholding was applied so that no value exceeded 1. We called this matrix *sampling matrix* and the value of each element determined the visibility of a given face feature through the feature background in a given video frame for this trial; more precisely, $\boldsymbol{p}_{ijk} = \boldsymbol{f}_{ik} \cdot s_{ijk} + b \cdot (1 - s_{ijk})$, where $\boldsymbol{p}_{ijk}$ are the pixel values to be displayed for face feature i on frame j in trial k, $\boldsymbol{f}_{ik}$ are the original pixel values of face feature i of the exemplar face selected for trial k, s_{ijk} is the sampling matrix value for face feature i on frame j in trial k, and b is the feature background color.

3.6.4 Experimental design

Each participant came to the laboratory twice and filled in a personal information questionnaire (education, age, sex, hours of sleep, alertness, concussion history, mental illness history, etc.) on the first session. Participants completed a total of 1000 sampled-face trials in each session; nine participants also completed in each session 100 additional whole-face trials in which a non-sampled exemplar face was shown for the same amount of time. Sampled-face and whole-face trials were randomly intermixed throughout the experiment. Each experimental session was divided in four equal-size blocks (of 250 or 275 trials) and blocks were interleaved with breaks of approximately 5 minutes. In addition, after every 5 trials, the screen automatically showed text indicating that the participants could take a few seconds to blink and rest their eyes before pressing a key to continue the experiment (participants were instructed not to blink during the trials themselves).

On each trial, a central fixation cross was shown to the participants for 1500 ms, after which the video stimulus appeared during 200 ms, superposed to the fixation cross, again followed by the fixation cross until the participant responded (the next trial then followed after an additional constant 1500 ms); a mid-gray background was always present. A fixed inter-trial interval was used so that participants could predict the onset of the trials. Half of the participants had to categorize the sex of the faces while the other half had to categorize their expression (happy or neutral). Participants had to respond as accurately and rapidly as possible with two keys on the keyboard (half of the participants had to use the opposite key combination from the other half, to counterbalance any motor effect).

3.6.5 Behavioral data analysis

One session from one participant was removed from all analyses because its mean accuracy was 50%; a session from a different participant was removed because of prominent EEG artifacts on a large subset of trials. Finally, one 275-trial block from still another participant was lost due to a technical error.

Accuracies and response times were z-scored within each 250- or 275-trial block. Trials with a z-scored response time below -3 or above 3, or with an absolute response time below 100 ms or above 2000 ms, were excluded from further analyses. Sampling matrices weighted by z-scored accuracies were then averaged together for each session. (Such a weighted sum is equivalent to a linear regression here since sampling was random.) Resulting *classification images* were averaged together within each subject and then within each task. Analyses were repeated with randomly permuted accuracies 10,000 times and a statistical threshold ($p < .05$, one-tailed, pixel level, corrected for familywise error rate (FWER)) was determined using the maximum statistic method (Holmes, Blair, Watson & Ford, 1996).

3.6.6 EEG data preprocessing

All preprocessing was performed with the help of functions from the Fieldtrip toolbox (Oostenveld, Fries, Maris & Schoffelen, 2011). EEG raw data from each session was

segmented in trials, filtered between 1 and 30 Hz with two successive 4th order Butterworth IIR filters, baseline corrected using the average activity between 500 ms and 250 ms before stimulus presentation, and downsampled to a 250 Hz sampling rate. Mastoid electrodes were removed due to poor signal-to-noise ratio on most subjects and data was re-referenced to an average reference. Anomalous trials, trials in which eye movements were occurring during the stimulus and anomalous electrodes were identified and removed following careful visual inspection of the data; bad channels were interpolated using a spherical spline. An ICA using Hyvärinen's fixed-point algorithm (Hyvärinen, 1999) was then performed to identify blink and eye movement artifacts. Bad components were identified and removed following careful visual inspection; between 1 and 5 (mean = 1.4) components were removed for each session. Finally, we computed single-trial current scalp density (CSD) waveforms using the spherical spline method (lambda = 1e-5, spline order = 4, degree of Legendre polynomials = 14; Kayser & Tenke, 2006; Tenke & Kayser, 2012); all further analyses were conducted on this CSD data.

3.6.7 EEG data analysis

In every experiment in which performance is not at ceiling level, part of the trials initially labeled as correct are correct only by chance: e.g., if 20% of responses are incorrect, this means that another 20% was in fact correct only by chance (since there is a 50% chance of being correct or incorrect when guessing). Here, we can verify which trials are comprised in this percentage of “falsely” correct trials by verifying which are the trials whose sampling matrices correlate the least to the behavioral classification image. Using this novel analysis method, we kept only true correct trials which were not correct merely by chance for further analyses.

Trials with a z-scored response time below -3 or above 3, or with an absolute response time below 100 ms or above 2000 ms, were excluded from the regression analyses. For each session, electrode and time point, regularized (ridge) multiple linear regressions were performed between the standardized feature × presentation time sampling planes and the standardized EEG amplitudes (Figure S1a). Resulting regression coefficients were convolved with a Gaussian kernel (standard deviation of 3 time points,

or 12 ms) in the EEG time dimension. Maps of regression coefficients were averaged within each subject and then across subjects within each task. Analyses were repeated with randomly permuted accuracies 1,000 times and statistical thresholds ($p < .05$, two-tailed, FWER-corrected) at both the pixel and cluster (2D clusters across EEG time and presentation time; using the summed cluster values; arbitrary primary threshold of $p < .01$, two-tailed, uncorrected) levels were determined using the maximum statistic method (Holmes et al., 1996). Analyses were restricted to time points between 30 ms and 600 ms from face onset. Results are displayed for representative PO7 (left occipito-temporal; LOT) and PO8 (right occipito-temporal; ROT) sensors but multiple comparison corrections were applied across all electrodes. Results were similar for most occipito-temporal sensors; data from all electrodes is available in an online repository (https://osf.io/3r782/).

To investigate whether processing was significantly modulated by the presentation moment and the task, a task × presentation moment ANOVA was performed. Maps of regression coefficients for each subject, face feature and electrode were first linearly interpolated to a resolution of 0.1 ms, realigned to the feature onset instead of the face onset (e.g., the EEG activity for the first presentation moment stayed the same, while activity for the second one was shifted left by 8.3 ms, activity for the third one by 16.7 ms, and so on), and resampled to the original resolution of 4 ms. Task × presentation moment ANOVAs were then performed on individual subjects' regression coefficients for each face feature, electrode, and latency from the feature onset (Figure S1b). Resulting F values were interpolated in topography space using biharmonic spline interpolation (Sandwell, 1987). Analyses were repeated on the 1,000 null maps obtained by randomly permuting accuracies and statistical thresholds ($p < .05$, one-tailed, FWER-corrected) at both the pixel and cluster (3D clusters across EEG time and topography space; using the summed cluster values; arbitrary primary threshold of $p < .01$, one-tailed, uncorrected) levels were determined using the maximum statistic method (Holmes et al., 1996). Analyses were restricted to time points between 50 ms and 400 ms from feature onset.

3.6.8 Mutual information between brain and behavior regression coefficients

For each subject, electrode and latency from feature onset, Gaussian copula mutual information (Ince et al., 2015; Ince, Giordano, Kayser, Rousselet, Gross & Schyns, 2017) was computed between the results of the behavior-stimulus weighted sum and the absolute values of the results of the EEG-stimulus regression, across reception moments (stimulus presentation time frames). Analyses were repeated with regression coefficients from the 1,000 null maps obtained by randomly permuting accuracies and statistical thresholds ($p <$ .05, one-tailed, FWER-corrected) at both the pixel and cluster (3D clusters across EEG time and topography space; using the summed cluster values; arbitrary primary threshold of $p < .01$, one-tailed, uncorrected) levels were determined using the maximum statistic method (Holmes et al., 1996). Analyses were restricted to time points between 50 ms and 400 ms from feature onset.

3.7 References

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459–1469.

Baluch, F. & Itti, L. Mechanisms of top-down attention. (2011). *Trends in Neuroscience. 34*, 210–224

Bartels, A. & Zeki, S. (2006). The temporal order of binding visual attributes. *Vision Research, 46*, 2280–2286.

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, *50*(12), 2830–2838.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Brisson, B., & Jolicoeur, P. (2007). The N2pc component and stimulus duration. *Neuroreport*, *18*(11), 1163–1166.

Broadbent, D. E. *Perception and communication.* (1958). Oxford, UK: Pergamon Press.

Caplette, L., Wicker, B., & Gosselin, F. (2016). Atypical Time Course of Object Recognition in Autism Spectrum Disorder. *Scientific Reports*, *6*:35494.

Caplette, L., Wicker, B., Gosselin, F., & West, G. (2017a). Hand position alters vision by modulating the time course of spatial frequency use. *Journal of Experimental Psychology: General*, *146*(7), 917–923.

Caplette, L., McCabe, E., Blais, C., & Gosselin, F. (2017b). The Time Course of Object, Scene and Face Categorization. In C. Lefebvre and H. Cohen (Eds.), *Handbook of Categorization in Cognitive Science* (2nd Edition). Amsterdam: Elsevier.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*, 1484–1525.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Di Russo, F., Stella, A., Spitoni, G., Strappini, F., Sdoia, S., Galati, G., et al. (2012). Spatiotemporal brain mapping of spatial attention effects on pattern-reversal ERPs. *Human Brain Mapping*, *33*(6), 1334–1351.

Dupuis-Roy, N., Faghel-Soubeyrand, S., & Gosselin, F. (2019). Time course of the use of chromatic and achromatic facial information for sex categorization. *Vision Research*, *157*, 36–43.

Dupuis-Roy, N., Fortin, I., Fiset, D., & Gosselin, F. (2009). Uncovering gender discrimination cues in a realistic setting. *Journal of Vision*, *9*(2):10.

Van Essen, D. C., Newsome, W. T. & Bixby, J. L. (1982). The pattern of interhemispheric connections and its relationship to extrastriate visual areas in the macaque monkey. *Journal of Neuroscience, 2*, 265–283.

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic Sampling within and between Objects despite Sustained Attention at a Cued Location. *Current Biology*, *23*(24), 2553–2558.

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, *22*(6), 1094–1118.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.

Holmes, A. P., Blair, R. C., Watson, J. D. G., & Ford, I. (1996). Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow and Metabolism*, *16*(1), 7–22.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, *8*(3), 197–230.

Hyvärinen, A. (1999). Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions: Neural Networks*, *10*, 626–634.

Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, *38*(3), 1541–1573.

Ince, R. A. A., Jaworska, K., Gross, J., Panzeri, S., van Rijsbergen, N. J., Rousselet, G. A., & Schyns, P. G. (2016). The Deceptively Simple N170 Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres. *Cerebral Cortex*, *26*(11), 4123–4135.

Ince, R. A. A., van Rijsbergen, N. J., Thut, G., Rousselet, G. A., Gross, J., Panzeri, S., & Schyns, P. G. (2015). Tracing the Flow of Perceptual Features in an Algorithmic Brain Network. *Scientific Reports*, *5*:17681.

Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. A dynamic causal modeling study. *Brain & Cognition*, *99*, 46–56.

Kayser, J. & Tenke, C. E. (2006). Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. evaluation with auditory oddball tasks. *Clinical Neurophysiology, 117*, 348–368.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, ECVP Abstract Supplement.

Landau, A. N., & Fries, P. (2012). Attention Samples Stimuli Rhythmically. *Current Biology*, *22*(11), 1000–1004.

Latour, P. L. (1967). Evidence of internal clocks in the human operator. *Acta Psychologica, 27,* 341–348.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* New York, USA: Henry Holt and Co.

Nagai, M., Bennett, P. J., & Sekuler, A. B. (2007). Spatiotemporal templates for detecting orientation-defined targets. *Journal of Vision*, *7*(8), 11–16.

Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience*, *5*(8), 812–816.

Neri, P., & Levi, D. M. (2007). Temporal Dynamics of Figure-Ground Segregation in Human Vision. *Journal of Neurophysiology*, *97*(1), 951–957.

Neri, P., & Levi, D. (2008). Temporal dynamics of directional selectivity in human vision. *Journal of Vision*, *8*(1):22.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735.

Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience, 2011*:156869.

Parker, D. M., Lishman, J. R., & Hughes, J. (1996). Role of coarse and fine spatial information in face and object processing. *Journal of Experimental Psychology. Human Perception and Performance*, *22*(6), 1448–1466.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Saenz, M. & Fine, I. (2010). Topographic organization of V1 projections through the corpus callosum in humains. *Neuroimage, 52*, 1224–1229.

Sandwell, D. T. (1987). Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data. *Geophysical Research Letters*, *14*, 139–142.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*(3), 243–265.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409.

Suzuki, M., & Noguchi, Y. (2013). Reversal of the face-inversion effect in N170 under unconscious visual processing. *Neuropsychologia*, *51*(3), 400–409.

Tanskanen, T., Näsänen, R., Ojanpää, H., & Hari, R. (2007). Face recognition and cortical responses: effect of stimulus duration. *Neuroimage*, *35*(4), 1636–1644.

Tenke, C. E. & Kayser, J. (2012). Generator localization by current source density (CSD): Implications of volume conduction and field closure at intracranial and scalp resolutions. *Clinical Neurophysiology*, *123*, 2328–2345.

Tse, P. U. (2004). Mapping visual attention with change blindness: new directions for a new method. *Cogitive Science*, *28,* 241–258.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643.

Ullman, S. (1984). Visual Routines. *Cognition*, *18*, 97–159.

Van Essen, D. C., Newsome, W. T. & Bixby, J. L. (1982). The pattern of interhemispheric connections and its relationship to extrastriate visual areas in the macaque monkey. *Journal of Neuroscience, 2*, 265–283.

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*:365.

Vinette, C., Gosselin, F., & Schyns, P. (2004). Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. *Cognitive Science*, *28*(2), 289–301.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A*, *4*(10), 2006–2021.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684.

Zhan, J., Ince, R. A. A., van Rijsbergen, N., & Schyns, P. G. (2019). Dynamic Construction of Reduced Representations in the Brain for Perceptual Decision Behavior. *Current Biology*, *29*, 319–326.

3.8 Supplementary Figures

Figure S1

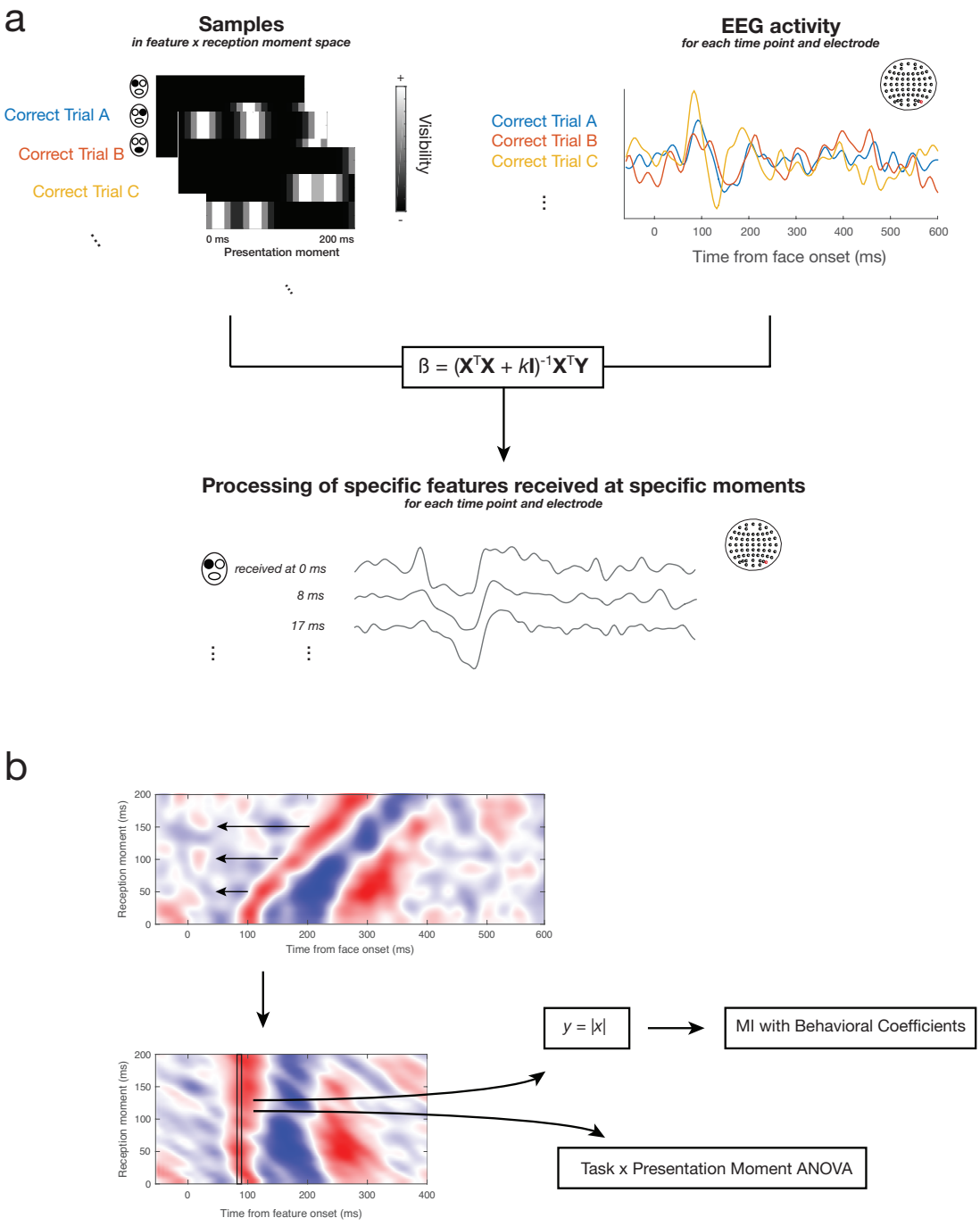


Figure S2

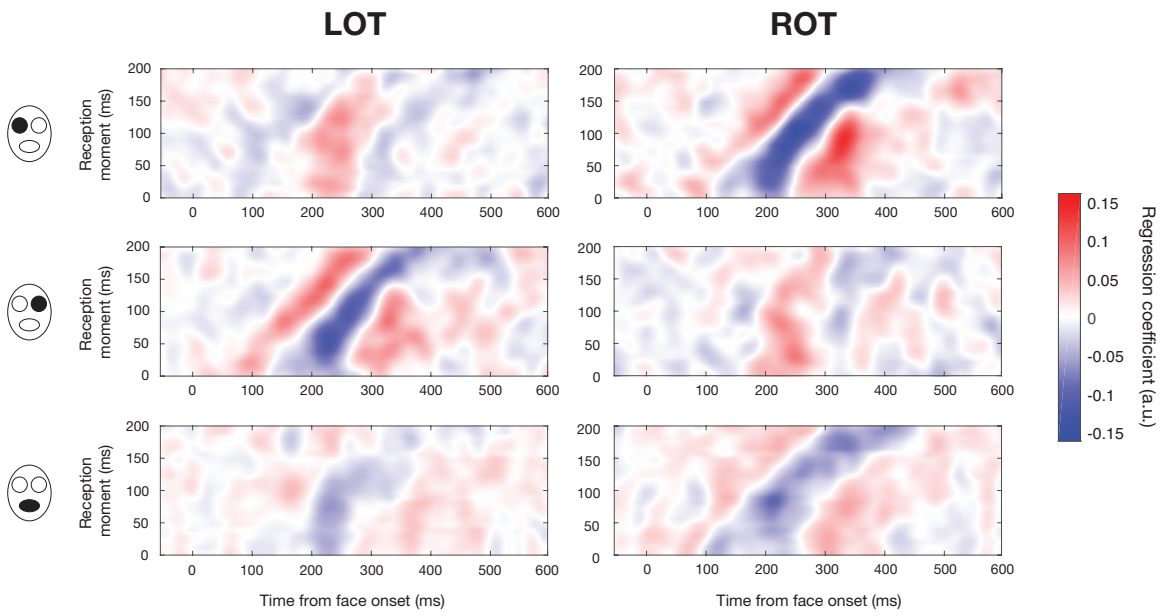
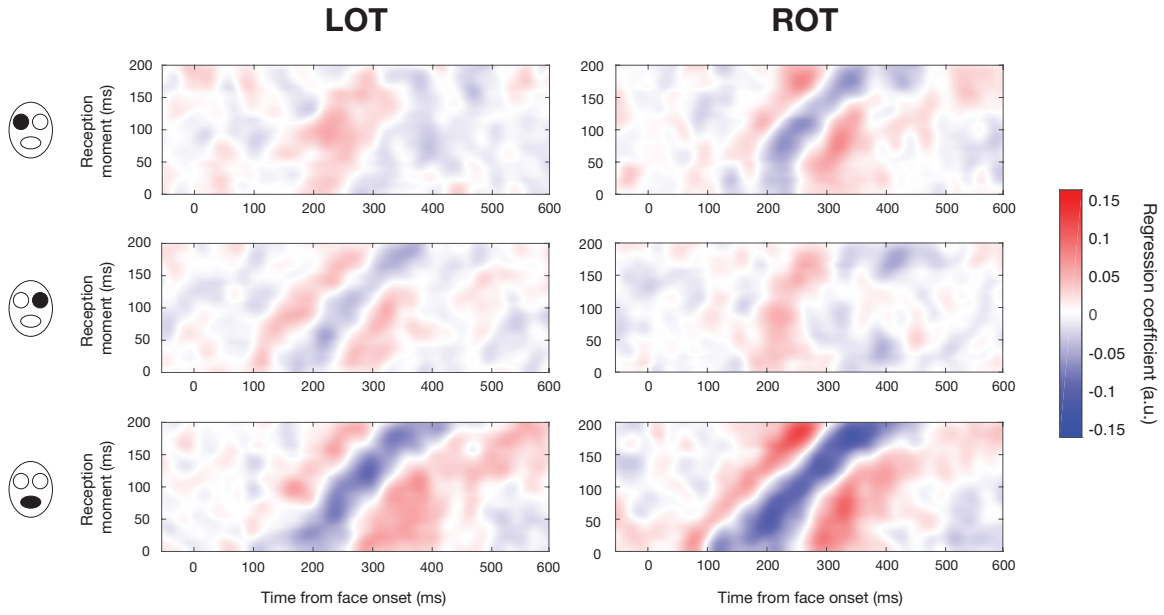


Figure S3



3.9 Supplementary Figure Legends

Figure S1. EEG data analyses. **A)** On each trial, a random sampling matrix determines how much each face feature is visible on each presentation moment (the *samples*). Only sampling matrices of truly correct trials (see Methods: EEG Data Analysis) are kept. On each corresponding trial, EEG activity is also recorded across the scalp for a certain period of time (examples are shown for one electrode). For each subject, samples (**X**; independent variable) and EEG activity (**Y**; dependent variable) are combined using a regularized (ridge) multiple linear regression, which allows us to uncover the EEG activity, across time and across the scalp (examples are shown for one electrode), related to the presentation of each specific face feature shown at each presentation moment. These time courses of regression coefficients can be arranged in images (maps) for specific face features and electrodes where amplitude is now represented by color (see panel B or figures 5 and 6 of the manuscript). **B)** Prior to further analyses, maps of regression coefficients are rearranged so that the zero point is the onset of the feature instead of the whole face (notice the change of the *x*-axis title). More specifically, EEG activity related to the presentation of a feature 8.3 ms after the face onset is shifted left by 8.3 ms, EEG activity related to the presentation of a feature 16.7 ms after the face onset is shifted left by 16.7 ms, etc. (see Methods: EEG Data Analysis). Only the first 400 ms are kept so that there is the same number of time points associated with each reception moment. Each 24-element column of this realigned image (activity across reception moments for each latency from the feature onset) is then submitted to subsequent analyses (example illustrated for one column). In the task x presentation moment ANOVA, columns are compared across subjects and the effect of the task (between-subject factor), the effect of the reception moment (within-subject factor), and the interaction between those factors are computed. Prior to the mutual information (MI) analysis, coefficients are transformed into their absolute values. For each subject, mutual information is then computed between the column of values and the vector of 24 values obtained in the behavioral analysis (see Methods: Behavioral data analysis) associated to the same face feature.

Figure S2. Mean maps of regression coefficients for the gender task, for LOT and ROT sensors, when including only trials in which there was one bubble (one feature revealed once).

Figure S3. Mean maps of regression coefficients for the expression task, for LOT and ROT sensors, when including only trials in which there was one bubble (one feature revealed once).

Chapitre 4

Article 3

Rhythmic sampling of visual features in the brain during object recognition

Laurent Caplette*, Karim Jerbi & Frédéric Gosselin

Department of Psychology, Université de Montréal, Montréal, Qc, Canada

Corresponding author: Laurent Caplette
Email: laurent.caplette@umontreal.ca
Address: Department of Psychology, University of Montreal, C.P. 6128 succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada

4.1 Abstract

During the fixation of an object, information is not only processed through time: it is also received on the retina through time. Because of several factors including ongoing brain oscillations, information is likely to be processed differently depending on when it is received. Here, we use magnetoencephalography (MEG) and a novel paradigm based on reverse correlation to reveal how specific visual features received at each moment during a fixation are processed through time and cortical space. We first show that information processing is highly dependent on when information is received: the same feature received at different moments can elicit large MEG responses or none at all. We further show that this sampling during the fixation is rhythmic: face features are sampled at multiple frequencies between 7 and 30 Hz across the parietal and occipital lobes. Finally, we show that different face features are largely sampled at distinct frequencies, demonstrating that frequency multiplexing is occurring at a large scale in the brain and that it is related to the sampling and not only to the processing of visual information.

4.2 Introduction

In addition to being processed through time in the brain, visual information is also received on the retina through time during the fixation of an object. This temporal aspect of perception is less discussed, but no less important (see VanRullen, 2011). Depending on the moment at which information is received, it may be processed differently or even not processed at all (Caplette, Ince, Jerbi & Gosselin, submitted). These variations in processing may be caused by several factors including the sampling of different features at different moments throughout fixation, and an overall rhythmic sampling of visual information, i.e. processing information more efficiently when it is received on the retina at specific moments reoccurring periodically (VanRullen, 2016).

Previous studies have demonstrated periodicities in perception and attention (see VanRullen, 2016, for a review). For instance, the visual threshold for detecting two successive flashes oscillates (around 33 Hz) as a function of the time interval between them (Latour, 1967), and the information used to recognize a face seems to be preferentially sampled at a frequency between 10 and 15 Hz (Blais, Arguin & Gosselin, 2013). In addition, when multiple stimuli are monitored simultaneously, attention seems to fluctuate between them at a 7-8 Hz frequency (Dugué, Roberts & Carrasco, 2016; Helfrich et al., 2018; Holcombe & Chen, 2013; Landau & Fries, 2012; Landau et al., 2015; Fiebelkorn, Saalman & Kastner, 2013; Fiebelkorn, Pinsk & Kastner, 2018; Re, Inbar, Richter & Landau, 2019; VanRullen, Carlson & Cavanagh, 2007); this rhythmic sampling also seems to occur when only one object is monitored (Holcombe & Chen, 2013; VanRullen et al., 2007). Relatedly, detection accuracy has been found to correlate with the phase of ongoing or prestimulus theta (~7 Hz) oscillations (Busch, Dubois & VanRullen, 2009; Busch & VanRullen, 2010; Hanslmayr, Volberg, Wimber, Dalal & Greenlee, 2013; Helfrich et al., 2018; Fiebelkorn et al., 2018), suggesting that this rhythmic perception is explained by underlying brain oscillations. Similarly, neural correlates of perception have been found to depend on prestimulus alpha (~10 Hz) or theta phase (Gruber et al., 2014; Hanslmayr et al., 2013; Jansen & Brandt, 1991).

This rhythmic attentional rhythm and its neural basis have been thoroughly investigated, but much less research has been dedicated to understanding how different

brain areas sample visual features through time, i.e. how they process information received at different moments during fixation. Different brain areas may sample information differently, in a way that is not necessarily correlated with behavior. Moreover, rhythmic sampling has been investigated mostly in a sustained attention context in which items are monitored for several seconds; however, objects in real life are typically fixated for 200-300 ms and visual recognition can occur with a stimulus duration inferior to 100 ms (e.g., Bacon-Macé, Macé, Fabre-Thorpe & Thorpe, 2005). In such a situation, multiple features cannot be sequentially sampled at a 7 Hz rhythm (one cycle = 142 ms).

Instead of a time-division multiplexing (TDM) in which different features are sequentially sampled at different phases of an oscillatory cycle (or similarly, at successive cycles of a higher frequency oscillation), a frequency-division multiplexing (FDM) in which different features are sampled at different frequencies might be used by the brain. FDM across theta and beta frequencies has been observed in the processing of visual features (Schyns, Thut & Gross, 2011; Smith, Gosselin & Schyns, 2006; Romei, Driver, Schyns & Thut, 2011; see Panzeri, Brunel, Logothetis & Kayser, 2010). If different features are processed at different frequencies, they may also be sampled from the world at different frequencies, as these two phenomena would be caused by the same underlying brain oscillations.

In the following experiment, we randomly revealed random parts of a face image at random moments across 200 ms (Figure 4.1) and recorded the brain activity of 5 subjects (over 5 sessions each) with MEG (see also Caplette et al., submitted). This allowed us to uncover how information received at specific moments during fixation is processed through time in different brain areas. By visualizing both sampling and processing dimensions simultaneously, we can furthermore distinguish between oscillations in sampling (i.e. rhythmic sampling) and oscillations in processing (i.e. stimulus-related brain oscillations), and examine the properties of sampling rhythms in different brain areas.

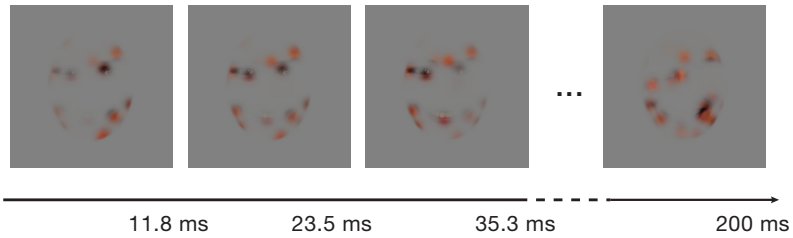


Figure 4.1. Illustration of an example stimulus in a random trial. Random areas of a random face image were smoothly revealed in random frames (1 frame each 11.8 ms) across 200 ms. See movies S1-S2.

4.3 Results and Discussion

4.3.1 Disentangling sampling and processing across the brain

For each subject and task, ridge regressions were performed between stimulus samples and MEG activity; to reduce the dimensionality of the data, spatial ROIs were constructed for the face features of interest (eyes and mouth) prior to the regression (see Methods; Figure 4.2a). Stimulus time × MEG time maps were thus obtained for each subject, task, face feature and source. These maps reveal an as-of-yet unseen dimension of visual processing: they show how information received on the retina at different moments throughout fixation is simultaneously processed through time throughout the brain. To identify different kinds of activation patterns across these maps without considering noisy maps with no signal, we used clustering by fast search and find of density peaks (Rodriguez & Laio, 2014). Example results from different clusters are displayed in Figure 4.2b.

We can observe on most maps a diagonal trend in the activations across stimulus moments: as expected, information received later is processed a similar amount of time later. But, on most sources, this processing is also remarkably variable depending on when it was received on the retina: information received at some moments elicits greater activity than information received at other moments (see all maps except bottom left, Figure 4.2b; Caplette et al., submitted). Moreover, different sources seem to preferentially process information presented and received at different moments (e.g., early vs late peak, top left vs middle left map, Figure 4.2b). We illustrate this variance across stimulus moments in Figure 4.3b,d. Large areas of the occipital and parietal lobes have significantly variable activity in all subjects ($p < .05$, one-tailed, FWER-corrected). It is worth noting that this

pattern of variance across the brain does not perfectly reflect the pattern of average signal across the brain (see mean activity in Figure 4.3a,c). For example, while there is generally similar or higher signal for the mouth compared to the left eye, there is a similar or higher variance for the left eye, especially in the gender task (Figure 4.3). In many cases, this variance seems to be caused by oscillations in sampling, that is, processed and unprocessed stimulus moments being alternated in an approximately regular manner (see Figure 4.2b, middle-middle and middle-bottom maps). We investigated this possibility next.

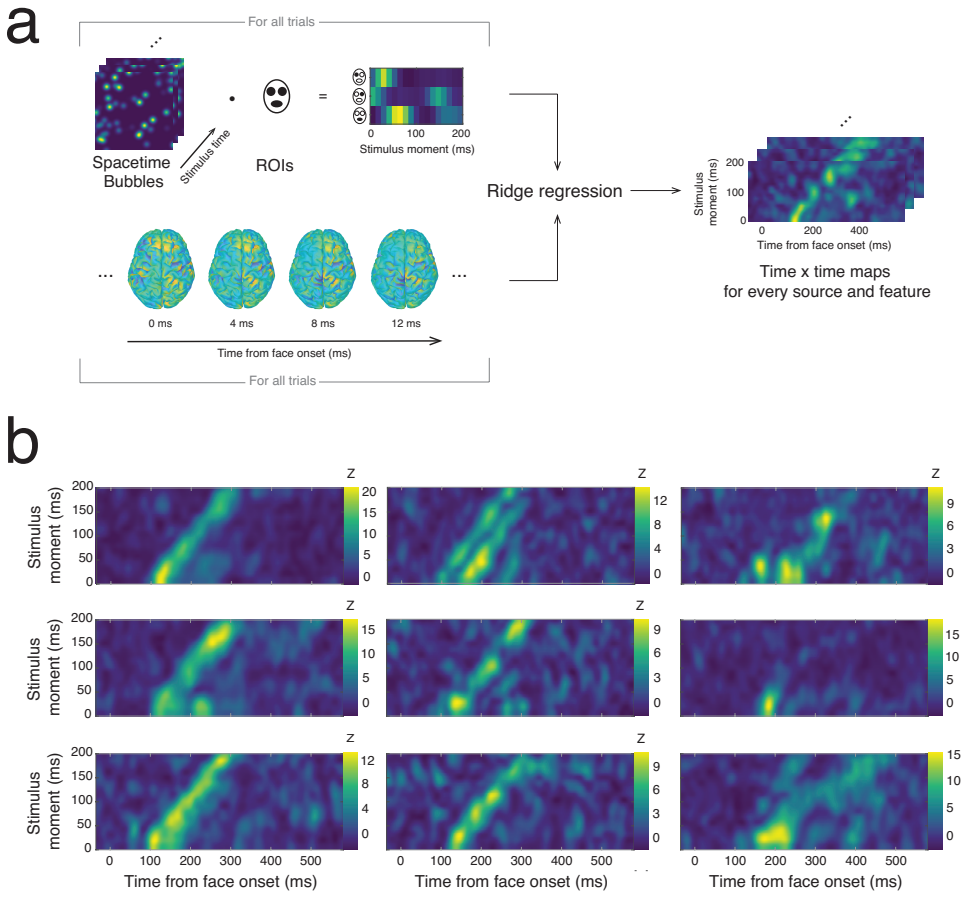


Figure 4.2. A) Illustration of the regression analysis for a given task and subject. For each trial, 3D space × time sampling matrices were dot multiplied with feature ROI masks and reduced to 2D feature × time samples: these constitute the independent variables of the multiple regression. MEG activity across trials for each source and time point constituted the dependent variables. A multiple ridge regression was performed for each dependent variable, and regression coefficients were rearranged in stimulus time × MEG time maps for each source and face feature. **B)** Example time × time maps for sources from different clusters as identified with density peak clustering. These maps originate from various subjects, tasks and face features. For each map, the *x* axis refers to the MEG time from the onset of the face stimulus; the *y* axis refers to the moment of presentation of the face feature within a trial, or *stimulus moment*.

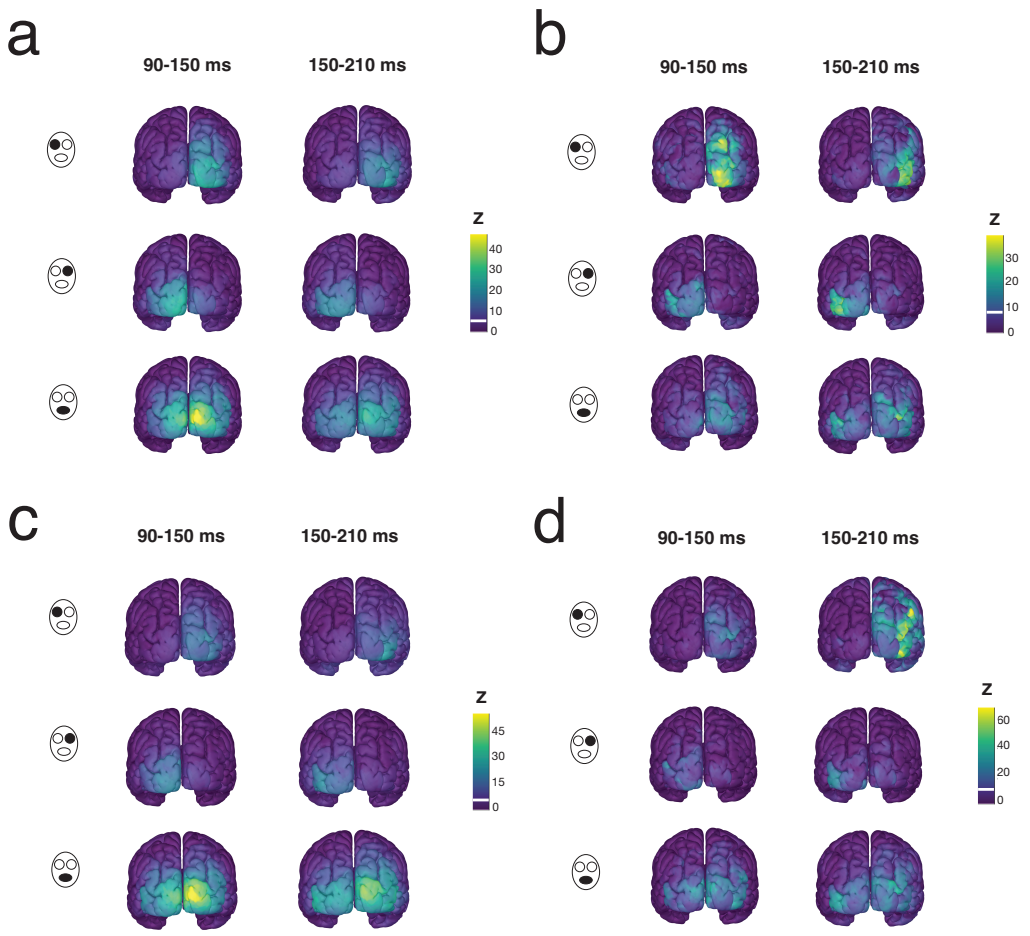


Figure 4.3. Activity mean (**A,C**) and variance (**B,D**) across stimulus moments, for each face feature, when averaging within two different time windows from the feature onset, for the gender (**A,B**) and expression (**C,D**) tasks. The white bar on the color legends indicates the statistical threshold ($p < .05$, one-tailed, FWER-corrected).

4.3.2 Oscillatory sampling across the brain

We performed a Fourier analysis of regression coefficients averaged between 90 to 150 ms from the feature onset, where most activity occurs (see Methods; Figure 4.4a). We uncovered significant oscillatory sampling between 6.8 and 30.4 Hz in all subjects and tasks, mostly in occipital areas (see group average in Figure 4.4b-c). Importantly, these results do not inform us about oscillations purely in processing, but about oscillations in the sampling of information across time during the fixation of a stimulus: that is, a significant oscillation here means that information was preferentially processed when it was received at specific moments occurring periodically.

This is, to our knowledge, the first direct demonstration of oscillatory sampling in the brain. While previous studies have shown that perception and attention are rhythmic, and have correlated these findings with oscillations in processing, they have not revealed how a specific brain area samples information from the world, and if it does so rhythmically. Since our focus was on visual object recognition during a typical 200 ms fixation, it is difficult to relate our results to the literature on rhythmic sustained attention (e.g., Busch & Vanrullen, 2010; Fiebelkorn et al., 2013; Landau & Fries, 2012). Despite this difference, similar oscillations in the theta and alpha range were uncovered (see Figure 4.4d). We also show significant low beta oscillations between 13 and 17 Hz: these frequencies are similar to those uncovered by Blais et al. (2013) who assessed the frequencies at which information was sampled by human subjects in a visual recognition task (see also Vinette, Gosselin & Schyns, 2004). Importantly, this oscillatory sampling reduces to oscillatory processing when we are looking at the processing of a feature irrespectively of when it was received on the retina (i.e. when collapsing across the *y* axis on the time × time maps). Decomposing processing in the way we did allows us to see that successive cycles of an underlying oscillation are allocated sequentially to information received at successive moments, instead of coding information received at the same moment. This is in fact probably the most logical way for the brain to proceed, since later cycles are fully dedicated to the most recent sensory information this way.

Notably, oscillatory sampling differed in frequencies between the different face features (Figure 4.4d). For example, for subject 1 in the gender task, the left eye is preferentially sampled at 6.8 and 16.9 Hz, while the right eye is preferentially sampled at 6.8 and 13.5 Hz and the mouth at 10.1 Hz. For subject 3, the spectra for the left eye, the right eye and the mouth peak respectively at 6.8, 10.1 and 20.3 Hz. For subject 5, all features are sampled at 6.8 Hz and the mouth is additionally sampled at 13.5 Hz. Overall, all subjects show a significant interaction between face features and frequencies in the

gender task. In the expression task, 3 out of the 5 subjects did not show any significantly oscillatory sources for features other than the mouth (see Methods for details about how oscillatory sources were determined). This is not really surprising, since the mouth is typically the feature most used to differentiate happy from neutral faces (e.g., Schyns et al.,

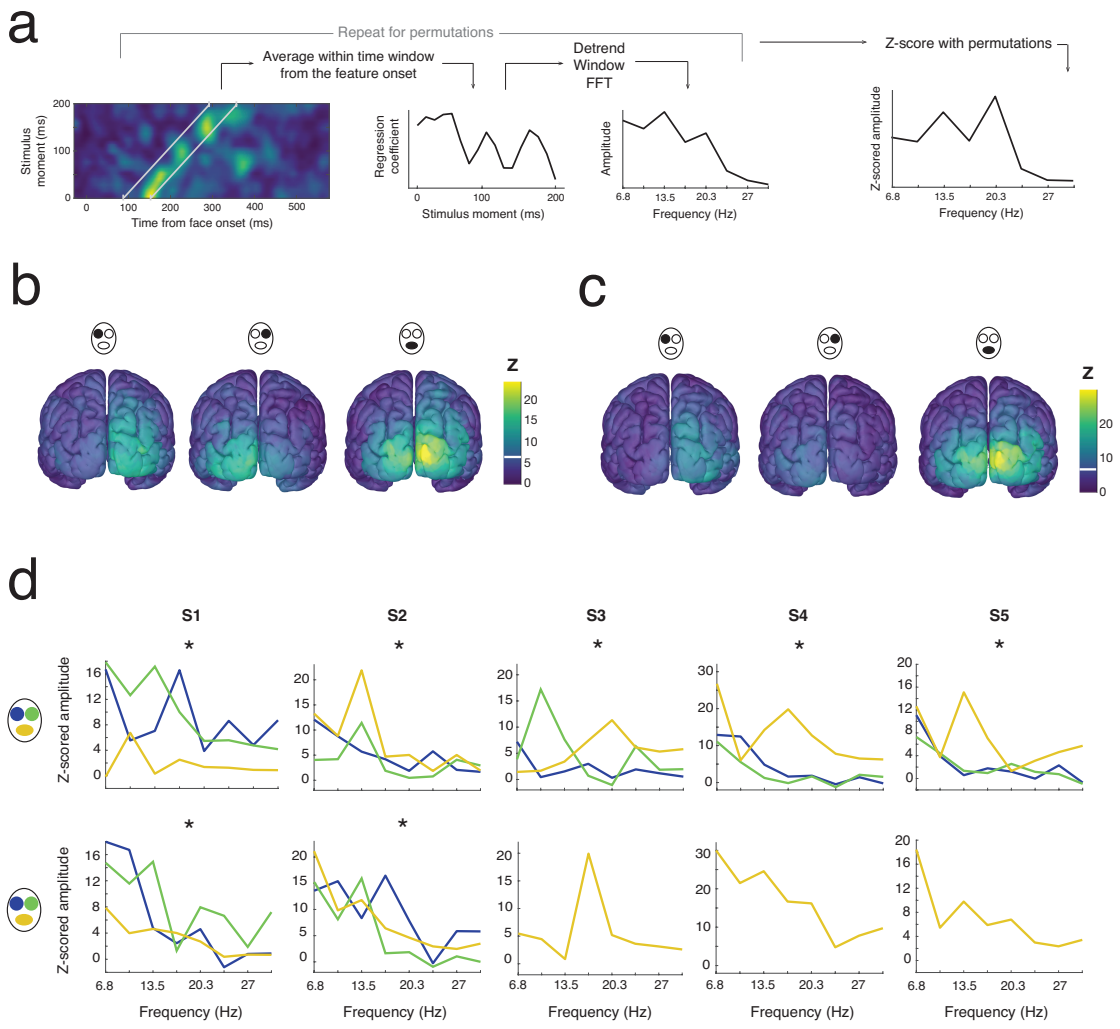


Figure 4.4. A) Illustration of the analysis of oscillatory sampling. For each subject, task, source and face feature, regression coefficients from the padded time × time maps (padding not illustrated) were averaged between 90 and 150 ms from feature onset (see diagonal gray lines on the leftmost image). This padded time series was linearly detrended, windowed and subjected to a Fast Fourier Transform (see Methods). The same analysis was repeated with the maps obtained from the permutation null distribution. Null distribution Fourier coefficients were used to z-score the observed Fourier coefficients; this normalization eliminated the 1/f background noise. **B)** Illustration, for each feature in the gender task, of the maximum oscillatory amplitude across frequencies, averaged across subjects. **C)** Same as B but for the expression task. **D)** Illustration of the average spectrum across oscillatory sources (see Methods) for each subject and face feature in the gender (upper row) and expression (lower row) tasks. Stars indicate significant interactions between face features and frequencies ($p < .05$, one-tailed, FWER-corrected). Note that for some subjects, there were oscillatory sources for only one face feature, preventing the test of an interaction.

2002). Since there was a spectrum for only one feature, this prevented the test of an interaction between face features and frequencies. Nevertheless, for the two other subjects, all three features were significantly oscillating and there was an interaction between features and frequencies.

These results are evidence of FDM in the sampling of information. Previous studies have shown that this type of multiplexing occurs during the processing of visual information in the parieto-occipital cortex. For example, Schyns et al. (2011) found that 4 Hz and 12 Hz EEG oscillations correlate with the presentation of the mouth and the eyes respectively; other studies also found that theta and beta frequency bands encoded different visual features during a visual recognition task (Smith et al., 2006; Romei et al., 2011). Furthermore, studies using single cell recordings observed that different low-level features such as contrast and orientation were coded on different timescales in the visual cortex (Bullock, 1997; Victor, 2000; see Panzeri et al., 2010). Importantly, these demonstrations of multiplexing regard the processing of information: these studies show how multiple features are being sent through the same channel by being coded at different frequencies. In our study, we show how different features are being sampled from the world at different frequencies. However, as discussed above, this reduces to (and explains) a multiplexed processing when collapsing across stimulus moments. Note that in our study, the different features seem to be coded at different frequencies in largely distinct areas. Arguably, this is not in itself multiplexing since different channels are used for different features. However, we must recall that these features need to be integrated together at some place in the brain: in all likelihood, this receiver channel would then need to perform FDM.

In summary, we showed how information received at specific moments during fixation is processed through time in different brain areas. In doing so, we revealed that information is processed in a significantly different way depending on when it is received on the retina and that it is in fact rhythmically sampled at multiple frequencies across the brain. We also showed that different features are sampled at different frequencies, in an instance of what is known as frequency-division multiplexing. This rhythmic sampling and multiplexing further illustrate that successive cycles of underlying brain oscillations are assigned to information received at successive moments instead of processing exactly the

same information, and thus that the brain processes the information most up to date from the sensory environment.

4.4 Methods

4.4.1 Participants

Three neurotypical adults (mean age = 28.3; SD = 3.2) were recruited on the campus of the University of Montreal; each came to the laboratory for a total of five experimental sessions. Participants did not suffer from any psychiatric or psychological disorder and had no known history of head concussions. The experimental protocol was approved by the ethics board of the Faculty of Arts and Sciences of the University of Montreal and the study was carried in accordance with the approved guidelines. Written informed consent was obtained from all the participants after the procedure had been fully explained, and a monetary compensation was provided upon completion of each experimental session.

4.4.2 Materials

The experimental program ran on a Dell Precision computer with Windows XP in the Matlab environment, using custom scripts and functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner, Brainard & Pelli, 2007). Stimuli were projected on a screen using a Sanyo PLC-XP41L projector, calibrated to allow a linear manipulation of luminance, with a resolution of 1152 x 864 pixels and an 85 Hz refresh rate. A viewing distance of 55 cm was maintained throughout the experiment. MEG activity was recorded using a 275-sensor CTF scanner with a sampling rate of 1200 Hz. Fiducials at the nasion, and left and right temple were used to track head movements in the scanner. Vertical electro-oculogram was bipolarly registered above and below the eyes, and horizontal electro-oculogram was bipolarly registered at the outer canthi of both eyes, to detect blinks and eye movements. An electrocardiogram was used to detect heartbeats. Prior to each session, the surface of the scalp and the positions of the fiducials were digitized using a Polhemus device.

4.4.3 Stimuli

Two hundred and sixty-four color images of faces were selected from the image database *Karolinska Directed Emotional Faces* (Goeleven, De Raedt, Leyman & Verschuere, 2008); only faces facing the camera were chosen. These were composed of 66 different identities (33 women and 33 men) each performing a happy and a neutral expression; two different pictures of each facial expression were used. Faces were aligned on twenty hand-annotated landmarks averaged to six mean coordinates for left and right eyes, left and right eyebrows, nose, and mouth, using a Procrustes transformation. Face images were then cropped and masked by a centered lightly smoothed elliptical mask (horizontal radius of 6 degrees of visual angle) to conceal the background, hair and shoulders. The mean luminance and contrast of all masked faces were equalized, separately for each color channel, using the SHINE toolbox (Willenbockel et al., 2010).

On each Bubbles trial, random areas of a randomly selected exemplar face were gradually revealed at random moments across a total duration of 200 ms (Caplette et al., submitted; Vinette et al., 2004). In other words, different random parts of a face were revealed on different frames of the stimulus, and the revealed areas were gradually appearing and disappearing (Figure 4.1; Movies S1-S2). A duration of 200 ms was chosen so that no saccade could occur during stimulus presentation on most trials. To create a Bubbles stimulus, we first generated a random 320 (pixels) × 320 (pixels) × 17 (frames) sparse matrix composed of zeros and a few ones, with some additional padding in all three dimensions. The probability of each element being one was adjusted on a trial-by-trial basis using a gradient descent algorithm to maintain accuracy around 75%. Three-dimensional Gaussian apertures (σ_{space} = 11.4 pixels; σ_{time} = 22.3 ms) were then centred on each one. On each trial, there was a minimum of one bubble in the stimulus. Superfluous padding was then removed, and thresholding was applied so that no value exceeded one. The matrix was finally dot-multiplied with the same elliptical face mask described above. We called this 3D matrix *sampling matrix* and the value of each element determined the visibility of a given pixel in a given frame for this trial. Specifically, the face image and the sampling matrix were dot-multiplied together, the complement of the sampling matrix was dot-

multiplied with a mid-gray plane and both results were summed together, so that non-sampled parts of the face were replaced by a median gray. Note also that on each trial, the underlying face image had a 50% probability of being flipped vertically, to compensate for possible informational differences between the left and right sides of the face images used.

4.4.4 Experimental design

Each participant came to the MEG Laboratory of the Department of Psychology, University of Montreal, five times. They filled a personal information questionnaire (education, age, sex, hours of sleep, alertness, concussion history, mental illness history, etc.) on the first session. Participants completed between 1000 and 1300 trials on each session. Bubbles trials were concatenated in blocks of 250 trials interleaved with breaks; MEG recording was stopped between blocks. On each block, participants performed either a Gender task ("man or woman?") or an Expression task ("happy or neutral?"). Tasks were alternated one block at a time and the first task performed was counterbalanced across subjects. In addition, participants were also shown whole non-sampled faces in 200 trials distributed across eight blocks (four of each task). These trials were randomly intermixed with Bubbles trials and participants had to perform the same task (faces were also shown for 200 ms). After every 5 trials, the screen automatically showed text indicating that the participants could take a few seconds to blink and rest their eyes before pressing a key to continue the experiment (participants were instructed not to blink during the trials themselves).

On each trial, a central black fixation cross was shown to the participants for 1500 ms, after which the video stimulus appeared during 200 ms, superposed to the fixation cross, again followed by the fixation cross until the participant responded (the next trial then followed after an additional constant 1500 ms); a mid-gray background was always present. A fixed inter-trial interval was used so that participants could predict the onset of the trials. Participants had to respond as accurately and rapidly as possible with two keys on the keyboard (key combinations were counterbalanced across participants).

4.4.5 MEG preprocessing and source reconstruction

Participants had all been MRI scanned in various studies over the course of the past five years; these anatomical T1 MRIs were used for source reconstruction in the current study. All MRIs were obtained using a MP-RAGE sequence on a 3T Siemens (Trio or Prisma) scanner with a 1 mm × 1 mm × 1 mm spacing.

The cortical surface was extracted using Freesurfer (http://surfer.nmr.mgh.harvard.edu) and downsampled to 8000 vertices using Brainstorm. All further preprocessing was conducted using functions from the Brainstorm toolbox (Tadel, Baillet, Mosher, Pantazis & Leahy, 2011). Anatomical landmarks corresponding to the locations of the MEG fiducials were identified and the MRI was aligned with the MEG sensors. The digitized surface of the scalp was used to refine the coregistration.

When necessary, MEG runs were subdivided so that the head did not move more than 3.5 mm within any run; short segments with large head movements were also excluded in the process. Data from all resulting runs was band-passed between 1 and 40 Hz using a Kaiser FIR Filter with 60 dB stopband attenuation and resampled to a 250 Hz sampling rate to reduce the dimensionality of the data. Bad channels were visually identified and removed. Signal Space Projection (SSP; Tesche et al., 1995) was used to identify and remove saccade, blink and heartbeat artifacts. The data was then segmented into trials from -250 ms to 600 ms from the stimulus onset, and baseline corrected using the average activity between 250 ms and 0 ms before the stimulus onset. Trials with noisy or anomalous segments, or with blinks/saccades during or close to the stimulus, were removed following a visual inspection.

For each run, a forward head model was constructed using overlapping spheres (Leahy, Mosher, Spencer, Huang & Lewine, 1998). A noise covariance matrix was then estimated from (filtered and resampled) empty-room noise recordings from the same day; it was regularized using automatic shrinkage during source estimation (Ledoit & Wolf, 2004). Estimation of the activity at all points of the cortical surface was performed for each trial using minimum norm imaging. One dipole, oriented normally to the cortical surface, was estimated at each vertex. Source maps were normalized using dynamical Statistical Parametric Mapping (dSPM; Dale et al., 2000).

4.4.6 MEG data analysis

In every experiment in which performance is not at ceiling level, part of the trials initially labeled as correct are correct only by chance: e.g., if 20% of responses are incorrect, this means that another 20% was in fact correct only by chance (since there is a 50% chance of being correct or incorrect when guessing). Here, we can verify which trials are comprised in this percentage of "falsely" correct trials by verifying which are the trials whose sampling matrices correlate the least to the behavioral classification image. Using this novel analysis method, we kept only true correct trials which were not correct merely by chance for further analyses (Caplette et al., submitted).

Response times were z-scored within each block; trials with a z-scored response time below -3 or above 3, or with an absolute response time below 100 ms or above 2000 ms, were excluded from further analyses. MEG activity was z-scored across (true correct) trials for each day, source and time point. Sampling matrices with their temporal padding were downsampled across the spatial dimensions and z-scored within each trial. ROIs for the three main face features (eyes and mouth) were then created using lightly smoothed circles for the eyes and a lightly smoothed ellipse for the mouth; all three ROIs contained a similar number of pixels (within a 1% margin). A scalar product was performed between the masks of the ROIs and the padded sampling matrices. These samples were z-scored across all true correct trials; they were then concatenated across all sessions, separately for each task. Source activity was also concatenated in the same way.

For each task, source, and time point, a ridge regression was performed between samples and source activity, using a regularization parameter of 20,000. Regression coefficients were then transformed into absolute values, and Gaussian-smoothed across the cortical surface (σ_{cortex} = 2.5 mm) and across time (σ_{time} = 12 ms). Analyses were repeated 250 times while randomly permuting sampling matrices across trials to establish an empirical null distribution. We then used clustering by fast search and find of density peaks (Rodriguez & Laio, 2014) to identify different kinds of activation patterns across these maps without considering noisy maps with no signal. We performed this clustering separately for each subject, task and face feature. Each time, between two and twelve

clusters were identified; exemplar maps from a sample of these were selected and shown in Figure 2b.

To assess the variance across stimulus moments, we averaged regression coefficients between 90 and 150 ms from the feature onset, and between 150 and 210 ms from the feature onset, for each stimulus moment, and computed the variance across the 17 stimulus frames. We repeated these analyses with the 250 permutation maps to obtain a null distribution of variance. Both the observed variance and the null distribution of variances were z-scored with the null distribution. A similar analysis was performed to assess the average activity across stimulus moments, computing the mean instead of the variance. We projected the results from individual subjects onto the MNI template brain, summed them and divided the result by the square root of 5, the number of participants (so that the results were still z-scores). We repeated this analysis with the null distribution and a statistical threshold ($p < .05$, one-tailed, pixel level, corrected for familywise error rate (FWER)) was determined using the maximum statistic method (Holmes, Blair, Watson & Ford, 1996).

4.4.7 Analysis of oscillatory sampling

For this analysis, regression coefficients were not smoothed across the cortical surface. Coefficients between 90 and 150 ms from the feature onset, where most brain activity occurs, were averaged together. For each source and face feature, resulting activity across all 25 (stimulus + padding) frames was linearly detrended, windowed and Fourier transformed. Given the short signals, we used a Tukey window with a taper value of 0.1. Fourier coefficients were then smoothed across the cortical surface (σ_{cortex} = 2.5 mm). These analyses were repeated on the null distribution regression coefficients. Both observed and null distribution Fourier coefficients were z-scored with the null distribution Fourier coefficients. Frequencies from 6.8 Hz to 30.4 Hz were analysed. To illustrate the presence of oscillations independently of the frequency, we computed the maximum amplitude across frequencies, for each subject, task and face feature. We then projected the results from individual subjects onto the MNI template brain, summed them and divided the result by the square root of 5 (see above). We repeated this analysis with the null

distribution and a statistical threshold ($p < .05$, one-tailed, pixel level, FWER-corrected) was determined using the maximum statistic method (Holmes et al., 1996).

We also computed the Fourier spectra for oscillatory sources for each subject, task and face feature. To do so, we performed again all previous analyses, including the regressions, but selecting only part of the trials. Specifically, we randomly divided the set of trials in half and performed all analyses twice (once for each half). We used one half to determine which sources were oscillatory for each face feature: they were deemed so if the maximum amplitude across frequencies exceeded the 5% significance level (FWER-corrected across features, frequencies, sources and tasks). We then used the other half of the data to compute the spectra on the sources previously deemed oscillatory (spectra were also computed for the null distribution of maps). We repeated this analysis while reversing the roles of the two data halves and averaged the two resulting spectra together. Both observed and null distribution Fourier coefficients were z-scored with the null distribution Fourier coefficients.

To determine whether there was a significant interaction between face features and frequencies, we computed for each subject and task the variance associated to face features, the variance associated to frequencies, and the variance associated to an interaction between face features and frequencies, using the z-scored spectra. Specifically, sums of squares were computed as in a regular two-way ANOVA, excluding the sum of squares associated to interindividual variability which was non-existent. We repeated this analysis with the z-scored null distribution spectra to establish a statistical threshold ($p < .05$, one-tailed, pixel level, FWER-corrected including across subjects) using the maximum statistic method (Holmes et al., 1996).

4.5 References

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459–1469.

Blais, C., Arguin, M., & Gosselin, F. (2013). Human visual processing oscillates: Evidence from a classification image technique. *Cognition*, *128*(3), 353–362.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Bullock, T. H. (1997). Signals and signs in the nervous system: The dynamic anatomy of electrical activity is probably information-rich. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(1), 1–6.

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37), 16048–16053.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience*, *29*(24), 7869–7876.

Caplette, L., Ince, R. A. A., Jerbi, K., & Gosselin, F. (submitted). Disentangling presentation and processing times in the brain.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron*, *26*(1), 55–67.

Dugué, L., Roberts, M., & Carrasco, M. (2016). Attention Reorients Periodically. *Current Biology*, *26*(12), 1595–1601.

Fiebelkorn, I. C., Pinsk, M. A., & Kastner, S. (2018). A Dynamic Interplay within the Frontoparietal Network Underlies Rhythmic Spatial Attention. *Neuron*, *99*(4), 842–853.

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic Sampling within and between Objects despite Sustained Attention at a Cued Location. *Current Biology*, *23*(24), 2553–2558.

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, *22*(6), 1094–1118.

Gruber, W. R., Zauner, A., Lechinger, J., Schabus, M., Kutil, R., & Klimesch, W. (2014). Alpha phase, temporal attention, and the generation of early event related potentials. *NeuroImage*, *103*(C), 119–129.

Hanslmayr, S., Volberg, G., Wimber, M., Dalal, S. S., & Greenlee, M. W. (2013). Prestimulus Oscillatory Phase at 7 Hz Gates Cortical Information Flow and Visual Perception. *Current Biology*, *23*(22), 2273–2278.

Helfrich, R. F., Fiebelkorn, I. C., Szczepanski, S. M., Lin, J. J., Parvizi, J., Knight, R. T., & Kastner, S. (2018). Neural Mechanisms of Sustained Attention Are Rhythmic. *Neuron*, *99*(4), 854–865.

Holcombe, A. O., & Chen, W. Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to <3 Hz when tracking three. *Journal of Vision*, *13*(1):12.

Holmes, A. P., Blair, R. C., Watson, J. D. G., & Ford, I. (1996). Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow and Metabolism*, *16*(1), 7–22.

Jansen, B. H., & Brandt, M. E. (1991). The effect of the phase of prestimulus alpha activity on the averaged visual evoked visual response. *Electroencephalography and Clinical Neurophysiology*, *81*(4), 241–250.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., et al. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2):e1005350.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, ECVP Abstract Supplement.

Landau, A. N., & Fries, P. (2012). Attention Samples Stimuli Rhythmically. *Current Biology*, *22*(11), 1000–1004.

Landau, A. N., Schreyer, H. M., van Pelt, S., & Fries, P. (2015). Distributed Attention Is Implemented through Theta- Rhythmic Gamma Modulation. *Current Biology*, *25*(17), 2332–2337.

Latour, P. L. (1967). Evidence of internal clocks in the human operator. *Acta Psychologica, 27,* 341–348.

Leahy, R. M., Mosher, J. C., Spencer, M. E., Huang, M. X., & Lewine, J. D. (1998). A study of dipole localization accuracy for MEG and EEG using a human skull phantom. *Electroencephalography and Clinical Neurophysiology*, *107*(2):159–173.

Ledoit, O., & Wolf, M. B. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*, 365–411.

Panzeri, S., Brunel, N., Logothetis, N. K., & Kayser, C. (2010). Sensory neural codes using multiplexed temporal scales. *Trends in Neurosciences*, *33*(3), 111–120.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Re, D., Inbar, M., Richter, C. G., & Landau, A. N. (2019). Feature-Based Attention Samples Stimuli Rhythmically. *Current Biology*, *29*(4), 693–699.

Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, *344*(6191), 1492–1496.

Romei, V., Driver, J., Schyns, P. G., & Thut, G. (2011). Rhythmic TMS over Parietal Cortex Links Distinct Brain Frequencies to Global versus Local Visual Processing. *Current Biology*, *21*(4), 334–337.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409.

Schyns, P. G., Thut, G., & Gross, J. (2011). Cracking the Code of Oscillatory Activity. *PLOS Biology*, *9*(5):e1001064.

Smith, M. L., Gosselin, F., & Schyns, P. G. (2006). Perceptual moments of conscious visual experience inferred from oscillatory brain activity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(14), 5626–5631.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*(3):879716.

Tesche, C. D., Uusitalo, M. A., Ilmoniemi, R. J., Huotilainen, M., Kajola, M., & Salonen, O. (1995). Signal-space projections of MEG data characterize both distributed and well-localized neuronal sources. *Electroencephalography and Clinical Neurophysiology*, *95*, 189–200.

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*:365.

VanRullen, R. (2016). Perceptual Cycles. *Trends in Cognitive Sciences*, *20*(10), 723–735.

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19204–19209.

Victor, J. D. (2000). How the brain uses time to represent and process visual information. *Brain Research*, *886*, 33–46.

Vinette, C., Gosselin, F., & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. *Cognitive Science*, *28*(2), 289–301.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684.

Chapitre 5

Discussion générale

5.1 Utilisation d'information visuelle à travers le temps

5.1.1 Avant-propos et définitions

Rappelons d'abord que par information *utilisée*, nous entendons l'information à la fois représentée et disponible dans l'input sensoriel (voir section 1.5.1). Ainsi, l'information utilisée à un moment précis ne consiste qu'en ce qui est à la fois représenté et disponible dans l'information reçue sur la rétine à ce moment-là. Si l'information disponible est en moyenne constante à travers le temps (comme c'est le cas dans les expériences discutées dans cette thèse, étant donné la constance des stimuli sous-jacents et l'échantillonnage temporel aléatoire), les variations que nous retrouvons à travers le temps sont nécessairement des variations dans la représentation[6]. Même s'il peut être tentant de concevoir les variations d'utilisation d'information à travers le temps comme un déplacement attentionnel, il faut garder en tête que des phénomènes tels que l'adaptation peuvent également être la cause de telles variations[7].

Nous nous concentrerons dans cette section sur l'utilisation comportementale, retrouvée en utilisant *Bubbles* avec une variable dépendante comportementale; nous nous attarderons à l'utilisation par des régions cérébrales précises à la section 5.2.

5.1.2 L'utilisation des fréquences spatiales lors de la reconnaissance d'objets

Toute image est décomposable, par une opération appelée analyse de Fourier, en une somme de grilles sinusoïdales de différentes orientations et fréquences spatiales (FS), chacune de ces grilles possédant une amplitude et une phase spécifiques. Les basses FS correspondent aux variations grossières de luminance dans une image (i.e. les contrastes

[6] Notons que le terme « représentation » tel qu'employé ici englobe toute l'activité mentale (plutôt qu'une hypothétique représentation statique provenant d'une aire de haut niveau, si une telle chose existe) : ainsi, la « représentation » mentale serait modifiée d'un instant à l'autre par de nombreux phénomènes de haut et de bas niveau. (Nous ne nous aventurerons pas ici dans le débat sur la définition et signification du terme « représentation » et référons plutôt le lecteur à Bobrow & Collins (1975) et Palmer (1978).)

[7] Évidemment, on peut aussi se demander ce que constitue en réalité un « déplacement attentionnel » mais, encore une fois, une discussion sur la définition d'« attention » dépasse le cadre de cette thèse (voir Di Lollo in press; James, 1890; Tsotsos, 2019).

globaux) et les hautes FS correspondent aux variations fines (i.e. les détails et les contours des objets). Dans l'article 1 de cette thèse, nous rapportons une utilisation de FS entre 0 et 32 cycles par image (cpi) environ (condition *no-expectation*; section 2.4.2). Nous allons maintenant brièvement résumer la littérature quant à l'utilisation des FS en reconnaissance d'objets, et y relier nos résultats.

Différentes bandes de FS ne sont pas utilisées de manière équivalente pour la reconnaissance d'objets. Non seulement les images naturelles ont-elles une énergie plus grande dans les basses FS que les plus hautes (elles ont typiquement un spectre de Fourier 1/f), mais les représentations des individus sont biaisées vers certaines bandes. En filtrant sélectivement des images en « basses » ou en « hautes » FS, diverses études rapportent des résultats contradictoires quant à l'importance des basses et des hautes FS (Calderone et al., 2013; Collin & McMullen, 2005; Harel & Bentin, 2009). Notons cependant que les seuils utilisés pour ces filtres varient grandement à travers les études (voir Caplette et al., 2014, pour une discussion). Lorsqu'une tâche de catégorisation au niveau de base (par opposition aux niveaux subordonné ou superordonné) était utilisée, ni les basses ni les hautes FS ne semblaient avoir un rôle privilégié (Collin & McMullen, 2005; Harel & Bentin, 2009). En utilisant *Bubbles* dans l'espace des FS avec des objets de la vie de tous les jours, Caplette et al. (2014) rapportent qu'en effet, c'est plutôt une bande de FS intermédiaire (ou du moins qui serait considérée ainsi par la plupart des études) qui est la plus utile pour l'identification d'objets. Plus précisément, des FS entre 14 et 24 cycles par objet (cpo; 16 à 28 cycles par image, cpi, ou 2.3 à 4 cycles par degré d'angle visuel, cpd) semblent être les plus utiles. L'information utilisée risque de varier selon la tâche et les stimuli employés (e.g., Schyns & Oliva, 1999).

En ce qui concerne l'utilisation des FS à travers le temps, plusieurs études ont observé une utilisation *coarse-to-fine*, c'est-à-dire une utilisation des basses FS précédant l'utilisation des hautes FS (Parker, Lishman & Hughes, 1992, 1997; Schyns & Oliva, 1994; Watt, 1987). Ces résultats ont été obtenus en présentant des images filtrées en basses et en hautes FS une à la suite de l'autre (Parker et al., 1992, 1997), en présentant des images en basses et en hautes FS pour différentes durées (Watt, 1987), ou en présentant des images hybrides (une image en basses FS et une autre en hautes FS superposées; Schyns & Oliva, 1994) pour différentes durées. Ces méthodes sont cependant assez grossières à la fois en

termes de résolution spatiale (i.e. les FS précises en jeu) et temporelle. En utilisant *Bubbles* dans les dimensions temporelle et des FS, Caplette et al. (2016) rapportent le décours temporel de l'utilisation des FS avec une plus grande précision : des « basses » FS de 1 à 20 cpi sont utilisées pendant toute la présentation du stimulus, alors que de plus hautes FS de 20 à 35 cpi sont utilisées dans le seconde moitié de la fixation. Ces résultats constituent la première évidence supportant une hypothèse d'Ullman (1984; voir aussi Caplette et al., 2017b) selon laquelle l'information globale ou les basses FS continuent à être échantillonnées pendant toute la reconnaissance, pour activer les représentations de nouveaux objets dans le champ visuel. Caplette et al. (2017a) rapportent également des résultats similaires dans leur expérience 1 : des FS de plus en plus hautes sont utilisées tout au long de la fixation, pendant que les FS les plus basses continuent d'être utilisées. Dans l'article 1 de cette thèse, nous n'avons malheureusement pas parfaitement répliqué les résultats de ces deux études (condition *no-expectation*). Malgré une utilisation continue des basses FS, les hautes FS semblent également être utilisées tout au long de la fixation (Figure 2.2a). On peut observer une augmentation du z-score moyen des hautes FS plus tard durant la fixation, mais cet effet est faible et non-significatif. Des causes possibles à cette non-réplication incluent le nombre d'essais par participant moins grand dans cette étude, des différences méthodologiques mineures entre les expériences et un lissage plus grand.

5.1.3 L'utilisation des attributs faciaux lors de la reconnaissance de visages

Les attributs faciaux utilisés pour la reconnaissance de visages dépendent de la tâche à effectuer (e.g., Schyns & Oliva, 1999). Dans l'article 2 de cette thèse, nous avons notamment révélé le décours temporel de l'utilisation des attributs faciaux dans une tâche de reconnaissance du sexe des visages. Des études précédentes utilisant diverses méthodes ont mis en évidence l'importance des yeux, des sourcils, de la mâchoire et du contour du visage (e.g., Brown & Perrett, 1993; Nestor & Tarr, 2008a, 2008b; Russell, 2003). En utilisant *Bubbles* dans l'espace photométrique × FS avec des visages en tons de gris (sans maquillage et avec les cheveux masqués), Schyns et al. (2002) ont observé que l'information relativement grossière (5.6–22.5 cycles par visage, cpv) au niveau des yeux et de la bouche corrélait significativement avec des réponses correctes lors d'une tâche de

discrimination du genre (voir aussi Gosselin & Schyns, 2001). En utilisant *Bubbles* dans l'espace photométrique avec des visages en couleurs (non contrôlés pour le maquillage et autres accessoires), Dupuis-Roy et al. (2009) ont quant à eux observé que les régions des yeux corrélaient avec des réponses correctes et que les régions des yeux et de la bouche corrélaient avec des réponses correctes et rapides. Dans un article récent avec une banque de visages similaire, Dupuis-Roy et al. (2019) ont utilisé *Bubbles* dans un espace photométrique × chromaticité × temps. Ils rapportent que l'information achromatique provenant de l'œil gauche est utilisée tout au long de la reconnaissance, que l'information achromatique provenant de l'œil droit est utilisée pendant la seconde moitié de la fixation et que l'information chromatique provenant de la région de la bouche est utilisée au début et à la fin de la fixation. Notons que, dans la plupart des études, les sujets semblent davantage utiliser les yeux que la bouche (Dupuis-Roy et al., 2009, 2019; Schyns et al., 2002). Dans notre étude, nous observons également une utilisation privilégiée des yeux par rapport à la bouche. De plus, l'utilisation des yeux est assez constante à travers le temps, mais les deux yeux (et particulièrement l'œil gauche) sont moins utilisés pour les tout premiers moments de présentations (Figure 3.3). Étrangement, l'œil droit semble plus utilisé que l'œil gauche dans notre cas — cet effet semble cependant être causé par un ou deux sujets sur 12 (cf. plus de 100 participants dans Dupuis-Roy et al., 2019). Malgré une méthode d'échantillonnage différente (échantillonnage de seulement les trois attributs faciaux principaux), nous retrouvons donc des résultats similaires aux études antérieures.

Nous rapportons également dans l'article 2 le décours temporel de l'utilisation des attributs faciaux dans une tâche de reconnaissance de l'expression du visage (souriant vs neutre). Quelques études ont investigué les attributs utilisés dans la reconnaissance de l'expression. Certaines études ont observé une utilisation plus grande du bas des visages par rapport au haut des visages (Dunlap, 1927; Ruckmick, 1921) alors que d'autres ne rapportent pas de différences entre différentes régions du visage (Baron-Cohen, Wheelwright & Jolliffe, 1997; Coleman, 1949; Frois-Wittman, 1930). En utilisant *Bubbles* dans l'espace photométrique, plusieurs rapportent une utilisation plus grande de la région de la bouche (e.g., Blais et al., 2012; Gosselin & Schyns, 2001; Schyns et al., 2002; Smith et al., 2005). Blais et al. (2012) rapportent également une utilisation significative (mais moins grande) de la région des yeux dans une tâche de catégorisation d'émotions (huit

alternatives : joie, tristesse, peur, colère, dégoût, douleur, surprise et neutralité). Les auteurs de cette dernière étude ont également étudié le décours temporel de l'utilisation; il s'agit à notre connaissance de la seule étude l'ayant fait. En utilisant une durée de présentation de 500 ms, ils rapportent une utilisation de la bouche en deux temps (vers 200 ms et 450 ms) et une utilisation des sourcils vers 350 ms; les plis nasolabiaux et les yeux sont légèrement utilisés vers 400-450 ms. En utilisant un échantillonnage des trois attributs principaux uniquement (article 2), nous avons observé une utilisation constante de la bouche. Malgré que nous ayons inclus les sourcils dans les régions des yeux, il se peut que l'absence de leur échantillonnage séparé nous ait empêché de retrouver leur utilisation significative, comme dans l'étude de Blais et al. (2012). Notons cependant que l'étude de Blais et al. (2012) a utilisé des expressions faciales dynamiques (i.e. un signal dynamique en plus d'un échantillonnage dynamique) : ainsi, ces résultats pourraient être dûs à une différence d'information à travers le temps et ils sont probablement peu comparables aux nôtres. Notre étude est la seule, à notre connaissance, à avoir évalué l'utilisation d'information à travers le temps dans le cadre d'une tâche de discrimination d'expression faciale statique. En conclusion, il appert que la bouche est de loin l'attribut le plus utile pour la catégorisation d'émotions, particulièrement si la tâche consiste à différencier un visage souriant d'un visage neutre. Finalement, notons que nous avons observé une utilisation légèrement moins élevée au début et à la fin de la fixation. Cet effet peut paraître curieux, ou encore susciter des doutes chez le lecteur quant à l'application correcte du *padding* dans la création des stimuli. Rassurons d'abord le lecteur que le *padding* a bien été appliqué et que cet effet n'est pas dû à son absence (si c'était le cas, cet effet serait également observé dans le cerveau et ce n'est pas le cas). Un effet similaire est d'ailleurs observé dans nos études précédentes évaluant la dimension temporelle (Caplette et al., 2016, 2017a). La raison exacte de cet effet demeure inconnue et demande davantage d'investigation, mais nous pouvons émettre l'hypothèse que cela pourrait être dû à une incertitude chez les sujets quant au moment exact du début (et de la fin) du stimulus à venir : les sujets ne se prépareraient ainsi à échantillonner l'information appropriée que dans les moments où ils sont relativement certains que le stimulus sera présenté.

5.2 Échantillonnage d'information à travers le cerveau

Dans les articles 2 et 3 de cette thèse, nous sommes allés plus loin et avons investigué l'utilisation d'information par des électrodes distinctes sur le scalp (article 2) ou par des sources distinctes dans le cerveau (article 3). Le traitement par le cerveau de l'information reçue à des moments spécifiques pendant la fixation n'a jamais été examiné auparavant. Dans ces deux articles, nous avons utilisé comme stimuli des visages. Ces stimuli ont l'avantage de tous être composés des mêmes attributs spatiaux : un échantillonnage spatial (ou spatiotemporel) est ainsi particulièrement approprié. Deux tâches de catégorisation binaire classiques ont été utilisées : la reconnaissance du sexe et la reconnaissance de l'expressivité (souriant ou neutre). Dans l'article 2, ces tâches sont effectuées par des participants différents; dans l'article 3, elles sont effectuées par les mêmes participants.

5.2.1 Variations temporelles dans l'échantillonnage

Dans l'article 2, nous avons effectué des régressions régularisées entre l'activité EEG sur le scalp (à chaque électrode et latence) et la présentation des trois attributs faciaux principaux à chaque 8,3 ms à travers 200 ms. Nous avons observé une modulation significative du traitement selon le moment de présentation (et de réception sur la rétine) de l'information visuelle. La plus marquante de ces modulations est sans aucun doute le délai de traitement supplémentaire des yeux sur les électrodes ipsilatérales pour les premiers moments de présentation, dans la tâche de reconnaissance du sexe (Figure 3.5). L'information présentée dans les premières 50 ms de fixation est en effet toute traitée approximativement au même moment, vers 220 ms après le début du stimulus; l'information présentée par la suite est quant à elle traitée environ 120 ms après qu'elle soit reçue, le moment de traitement augmentant linéairement selon le moment de présentation, tel qu'on pourrait s'y attendre. Une explication possible à cet événement est l'existence d'un blocage du transfert d'information de l'hémisphère contralatéral à l'hémisphère ipsilatéral, autour de 220 ms. L'événement responsable de ce blocage pourrait être la N170 : la latence de 220 ms est compatible avec la fin de ce potentiel évoqué relié à la reconnaissance de visages et la N170 a déjà été associée au transfert interhémisphérique

des yeux lors de la reconnaissance de visages (Ince et al., 2016; voir la discussion de l'article 2). D'autres modulations significatives sont également rapportées : par exemple, la présence de deux inflexions négatives successives (plutôt qu'une seule) dans la représentation contralatérale des yeux dans la tâche de reconnaissance du sexe. Ces inflexions pourraient correspondre à des générateurs distincts de la N170 (Di Russo et al., 2012; Suzuki & Noguchi, 2013). Finalement, des différences de latences sont visibles à travers les moments de présentation. Notamment, la latence de la dernière composante positive (analogue à la P3) est plus courte pour les moments de présentation plus tardifs. Il semble donc qu'il y ait une certaine compression temporelle : l'information reçue plus tôt est maintenue pour davantage de temps que l'information reçue plus tard. Cela pourrait indiquer un processus d'accumulation. Ces résultats sont compatibles avec d'autres études rapportant des liens entre une composante à des latences similaires et l'accumulation/intégration temporelle d'information (O'Connell, Dockree & Kelly, 2012; Twomey et al., 2015).

Notons que ces modulations ne peuvent être causées par de l'adaptation ou un amorçage, étant donné que des résultats similaires sont obtenus lorsque seuls les essais avec une seule bulle sont inclus. Ces différences selon le moment de présentation semblent également être d'origine au moins partiellement descendante, puisqu'elles sont modulées par la tâche. De plus, elles sont reliées aux modulations de l'utilisation comportementale à travers le temps.

Dans l'article 3, nous avons effectué des régressions régularisées entre l'activité MEG à chaque source et latence et la présentation des trois attributs faciaux principaux à chaque 11,8 ms à travers 200 ms. Nous avons d'abord observé chez tous les sujets une modulation hautement significative du traitement selon le moment de présentation sur une large partie des lobes occipitaux et pariétaux (et même sur le lobe frontal gauche dans le cas du sujet 1). Lorsqu'on regarde de plus près, on observe que cette modulation semble dans certains cas être causée par une variation linéaire (typiquement une plus grande activation pour les premiers moments de présentation), mais dans d'autres cas par une variation non-linéaire potentiellement oscillatoire où certains moments sont davantage traités que d'autres.

5.2.2 Échantillonnage rythmique

Dans l'article 3, nous avons formellement vérifié s'il y avait un échantillonnage oscillatoire dans le traitement ayant lieu entre 90 et 150 ms après la présentation de l'information. Nous avons observé chez tous les sujets un échantillonnage entre des fréquences de 6.8 et 23.7 Hz (résolution de 3.4 Hz), en plus de quelques sources à une fréquence de 30.4 Hz chez le sujet 1. Typiquement, plusieurs fréquences étaient significatives pour une source donnée pour le traitement d'un attribut donné. Il s'agit à notre connaissance de la première démonstration d'un échantillonnage rythmique dans le cerveau. En effet, plusieurs études ont démontré comportementalement un échantillonnage attentionnel rythmique et l'ont relié à des oscillations cérébrales (par exemple en reliant la phase des oscillations dans une région donnée au taux de réponses correctes) mais aucune ne semble avoir investigué comment une région cérébrale traite l'information reçue à des moments spécifiques tout au long de la fixation et si un tel échantillonnage est rythmique. Il est possible que différentes régions échantillonnent l'information à différentes fréquences et que ces oscillations ne soient pas nécessairement corrélées à l'échantillonnage comportemental.

Si l'on n'avait pas examiné le traitement de l'information reçue à des moments spécifiques, l'échantillonnage rythmique observé dans notre étude serait apparent comme une oscillation dans le traitement. Cela est visible lorsqu'on observe une carte temps × temps dans laquelle il y a un échantillonnage oscillatoire (e.g., Figure 4.2a ou Figure 4.2c, sujet 1) et qu'on moyenne l'activité à travers la dimension de la présentation du stimulus. En effet, si on observe, disons, un traitement de l'information reçue à 0 ms à une latence de 100 ms, aucun traitement de l'information reçue à 50 ms, un traitement de l'information reçue à 100 ms à une latence de 200 ms, aucun traitement de l'information reçue à 150 ms, et un traitement de l'information reçue à 200 ms à une latence de 300 ms, cela signifie que le traitement de cette information, lorsque l'on n'a pas accès à la dimension de la présentation du stimulus, survient à des latences de 100, 200 et 300 ms — en d'autres mots, qu'il oscille à une fréquence de 10 Hz. Ainsi, nos résultats nous permettent de constater que les cycles successifs d'une oscillation cérébrale sous-jacente sont alloués à l'information reçue à des moments successifs (Figure 5.1b). Il aurait aussi été possible que l'information reçue continue à être traitée indéfiniment selon cette oscillation (Figure 5.1c)

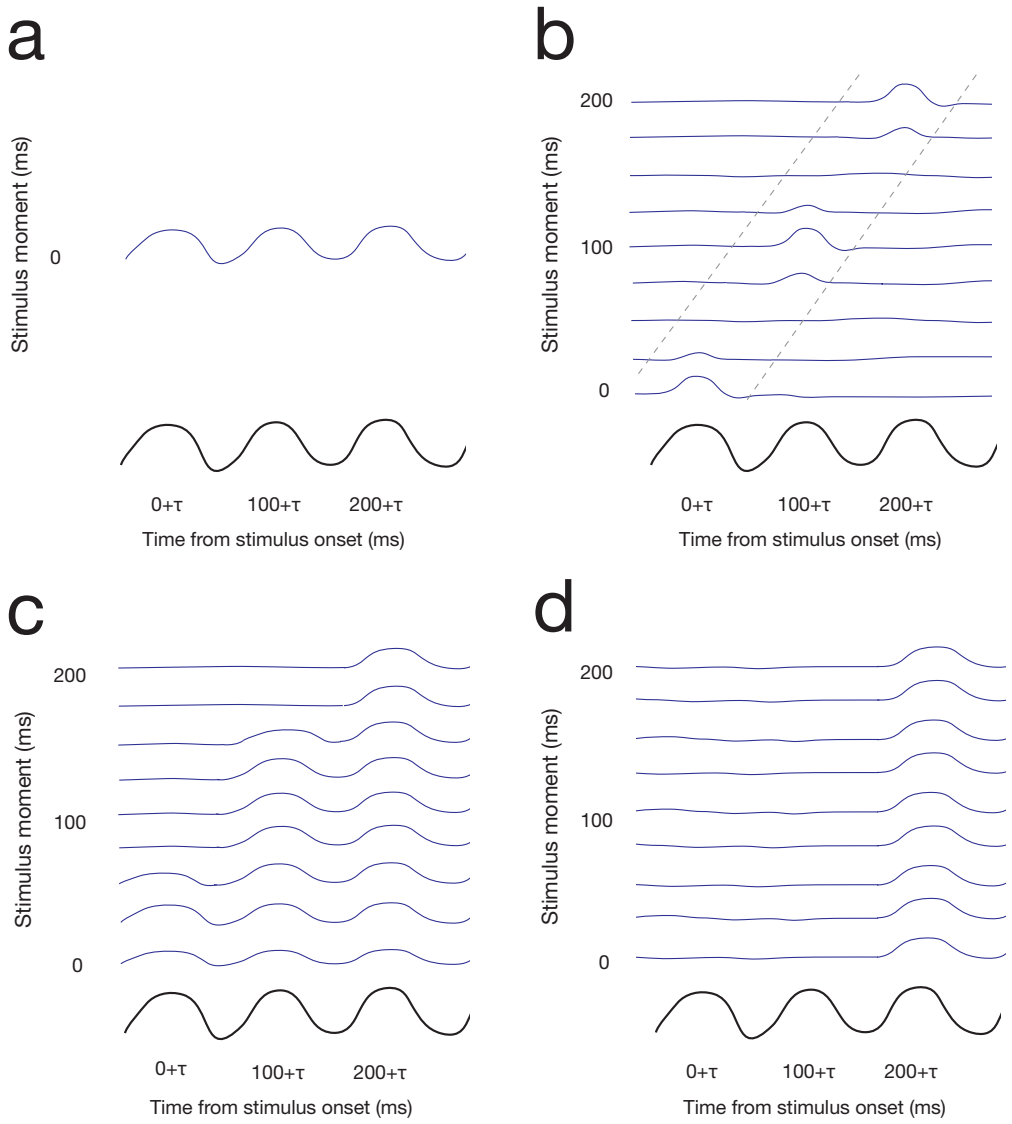


Figure 5.1. Quelques modulations possibles du traitement de l'information visuelle par une oscillation cérébrale sous-jacente. Sur chaque panneau, l'oscillation en noir au bas du panneau représente l'oscillation cérébrale sous-jacente dans une région cérébrale donnée (ici à 10 Hz), l'abscisse représente le temps depuis le début de la présentation d'un stimulus (τ désigne le délai que prend l'information visuelle pour parvenir à cette région) et l'ordonnée représente le moment de réception sur la rétine de l'information visuelle (notons que les deux axes réfèrent au même temps; 0 sur les deux axes réfère ainsi au même instant). **A)** Le traitement oscillatoire d'un stimulus si le stimulus est considéré comme un événement instantané, ce qui est un postulat implicitement accepté par plusieurs études. L'oscillation sous-jacente module directement le traitement du stimulus. **B)** Le traitement observé dans notre étude dans certaines régions cérébrales. Un traitement diagonal oscillatoire est visible (voir lignes pointillées), où l'information est traitée à certains moments seulement (survenant périodiquement) et où elle est traitée à une latence fixe après la réception de l'information lorsqu'elle est traitée. De plus, l'information n'est traitée que momentanément seulement. Un tel traitement peut être le résultat de l'information étant seulement traitée lorsqu'elle arrive dans la région cérébrale au moment où l'oscillation sous-jacente est dans une « bonne » phase et que ce traitement ne dure que jusqu'à ce que l'oscillation ne soit plus dans une bonne phase, les cycles subséquents de l'oscillation étant plutôt consacrés à l'information nouvellement arrivée. Si le traitement n'était pas décomposé selon le moment d'arrivée de l'information sur la rétine, ce traitement serait visible comme une simple oscillation dans le temps. **C)** Une autre possibilité expliquant un traitement oscillatoire : l'information est traitée de manière oscillatoire, conformément à la phase de l'oscillation cérébrale sous-jacente, mais elle continue d'être traitée indéfiniment selon cette oscillation. **D)** Ici, l'information est traitée de manière oscillatoire (l'oscillation continuant hors du cadre de l'image), mais cette oscillation débute seulement une fois que toute l'information est reçue et prête à être traitée. Une telle possibilité suppose une accumulation préalable de toute cette information (dans une région précédente).

ou encore que l'oscillation dans le traitement ne débute qu'une fois toute l'information reçue (Figure 5.1d), entre autres possibilités. Le postulat implicite que font plusieurs études de considérer les stimuli comme des événements instantanés ne permet pas de considérer toutes ces possibilités (Figure 5.1a). Le traitement que nous avons découvert consiste probablement en la manière la plus logique pour le cerveau de représenter l'information, puisque les cycles les plus récents de l'oscillation sont ainsi consacrés complètement au traitement de l'information sensorielle la plus récente (et donc la plus susceptible de nécessiter un changement de comportement).

Notons que pour observer les possibilités représentées à la Figure 5.1 avec une méthode telle que la nôtre, qui moyenne à travers plusieurs essais, nous devons supposer que la phase des oscillations sous-jacentes est réinitialisée et que c'est ainsi environ la même dans la plupart des essais. Il s'agit là d'un postulat raisonnable que beaucoup d'études précédentes sur l'attention (e.g., Fiebelkorn et al., 2013, 2018; Helfrich et al., 2018; Landau & Fries, 2012; Landau, Schreyer, Van Pelt & Fries, 2015) ont adopté. De plus, une analyse sommaire des phases par essai démontre que c'est largement le cas (données non montrées). Notons également que nous avons ici fait en sorte que les stimuli commencent toujours à un moment prédictible pour les sujets, au cas où cela serait important.

Un problème qu'on peut soulever avec la théorie que nous proposons (Figure 5.1b) est que l'information présentée au début de la fixation pourrait être très pertinente (par exemple dans un cas où les stimuli seraient dynamiques et qu'une information cruciale est présentée au début de la fixation) et qu'il faut donc s'en souvenir. Notons que notre modèle n'empêche pas une telle chose : l'information reçue à chaque moment est simplement transmise, aussitôt traitée, à une autre région qui intègre l'information reçue à chaque moment (voir section 5.3).

5.2.3 Multiplexage fréquentiel

En analysant les fréquences maximales dans les sources avec une activité oscillatoire significative, nous avons également observé que les différents attributs sont échantillonnés

maximalement à différentes fréquences (voir article 3); ces fréquences sont de plus différentes selon la tâche, pour le même attribut.

Le fait que les différents attributs sont traités à différentes fréquences nous indique qu'un multiplexage fréquentiel a lieu dans le cerveau. Plusieurs études avaient observé un tel multiplexage (Bullock, 1997; Schyns et al., 2011; Smith et al., 2006; Romei et al., 2011; Victor, 2000; voir Panzeri, Brunel, Logothetis & Kayser, 2010). Ces observations étaient toutefois encore une fois au niveau du traitement : différents attributs étaient traités à différentes fréquences. Ici, le multiplexage a lieu sur la dimension de l'échantillonnage (voir également Dupuis-Roy, 2014, article 3) : différents attributs sont échantillonnés du monde extérieur à différentes fréquences. Comme discuté plus haut (section 5.2.2), ces oscillations dans l'échantillonnage se réduisent à des oscillations dans le traitement si l'on ne tient pas compte du moment de réception de l'information. Ainsi, il semble que les oscillations cérébrales sous-jacentes expliquant l'échantillonnage et le traitement rythmique ne sont pas aux mêmes fréquences selon l'attribut en question. Pourquoi et comment différentes fréquences sont attribuées à différents attributs est une question d'intérêt pour de futures investigations.

Notons que dans notre étude, les oscillations aux différentes fréquences avaient lieu dans différentes régions, ce qui ne correspond pas parfaitement à la définition classique du multiplexage puisque différents canaux sont utilisés pour véhiculer les différentes informations (voir cependant le codage de l'œil gauche et de la bouche à différentes fréquences dans les mêmes sources chez le sujet 1; Figure 4.3c). Toutefois, notons qu'il s'agit probablement là d'une conséquence de notre échantillonnage spatial : le fait que les attributs soient situés à différents endroits du champ visuel implique qu'ils risquent d'être codés par des populations de neurones entièrement différentes. De plus, même si ces représentations initiales ont lieu dans différentes régions, ces différents attributs doivent être intégrés à un certain endroit dans le cerveau : il est probable que certaines régions doivent effectuer un multiplexage fréquentiel pour représenter simultanément ces différents attributs (voir sections 5.3 et 5.6). Néanmoins, multiplexage ou non, le codage de ces différents attributs à différentes fréquences demeure intéressant.

5.2.4 Comparaison des résultats EEG et MEG

Nous avons observé des variations significatives dans l'échantillonnage à la fois dans l'article 2 et dans l'article 3. De plus, l'activation autour de 100 ms après la présentation d'un attribut semble avoir une plus grande amplitude pour l'information reçue plus tard dans les deux cas (voir les images temps × temps pour le sujet 1, rangées du haut et du bas, et pour le sujet 3, rangée du bas, dans la Figure 4.2b; notez que les autres images de la figure 4.2b rapportant une activation plus grande pour l'information reçue tôt sont des sources avec une activation plus tardive autour de 150-160 ms et donc que ces résultats ne s'appliquent pas à notre propos). L'activation plus tardive autour de 300 ms est également plus grande pour l'information reçue tôt dans les deux articles (données non montrées). (Notons que dans l'article 3 qui évalue l'activité sur différentes sources avec précision, les différentes activations plus ou moins tardives sont en grande partie sur différentes sources, alors que dans l'article 2 qui évalue l'activité sur le scalp, les différentes activations sont visibles sur les mêmes électrodes.)

Plusieurs différences notables sont également présentes entre les deux études. D'abord, des oscillations sont clairement visibles sur plusieurs sources dans l'article 3 et leur amplitude est significative, alors qu'elles ne sont pas visibles dans l'article 2. Notons cependant qu'une certaine activité rythmique est présente dans l'article 2. Nous le démontrons dans une analyse supplémentaire qui n'est pas présentée dans l'article (Figure 5.2). Ces oscillations se limitent cependant à des fréquences plus petites et à une amplitude beaucoup plus faible, et aucun multiplexage fréquentiel n'est détectable. Une autre différence notable entre les deux études consiste en l'absence apparente (une analyse exhaustive de toutes les sources n'a toutefois pas été effectuée) dans l'article 3 d'un effet de blocage pour les yeux dans l'hémisphère ipsilatéral.

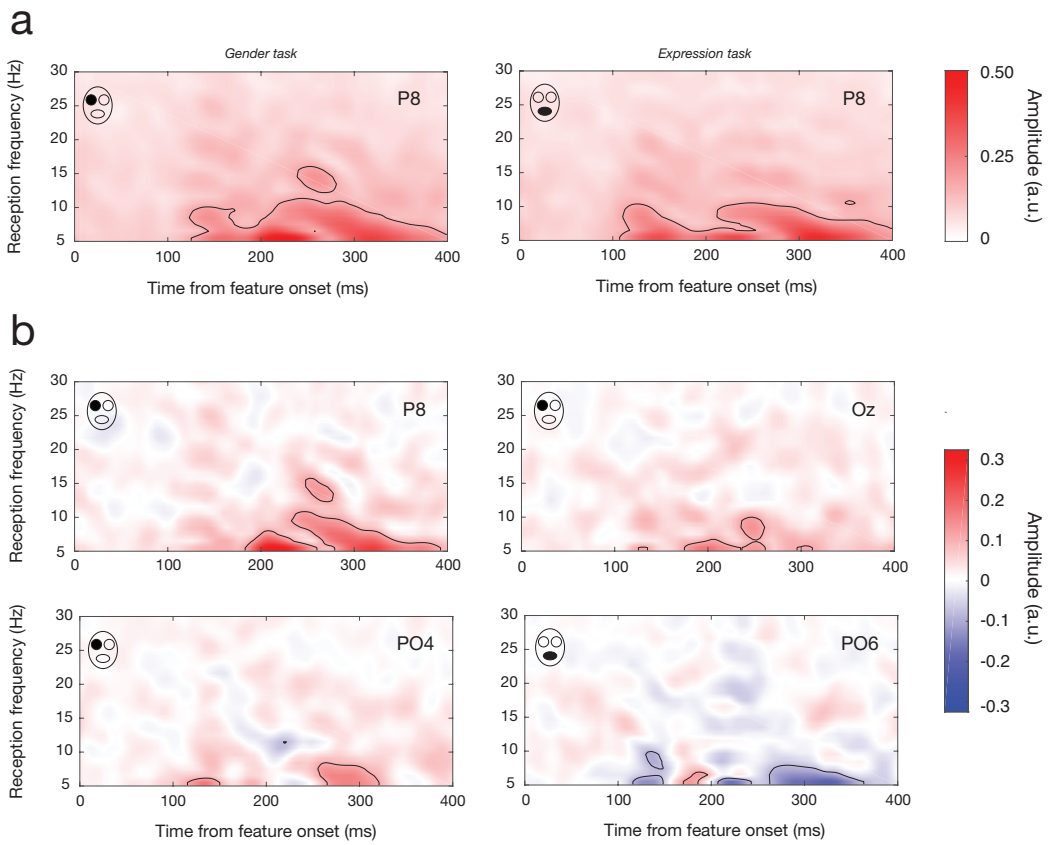


Figure 5.2. Illustration des oscillations dans l'échantillonnage dans l'article 2. **A)** Amplitude des oscillations pour l'œil gauche dans la tâche de reconnaissance du sexe et pour la bouche dans la tâche de reconnaissance d'expression, sur l'électrode P8. **B)** Illustration de la différence entre les deux tâches pour ce qui est de l'amplitude des oscillations, pour différents attributs et électrodes.

Plusieurs différences méthodologiques entre les expériences pourraient être à l'origine des différences dans les résultats : nous discutons ici des trois plus importantes. D'abord, la différence la plus évidente : l'article 2 est en EEG et rapporte l'activité sur le scalp, alors que l'article 3 est en MEG et rapporte l'activité dans le cerveau. Il est possible que le moyennage dans l'article 2 de l'activité de plusieurs parties du cerveau qui oscillent à différentes fréquences et phases ait éliminé l'échantillonnage rythmique dans les cartes temps × temps et que seules les variations non-oscillatoires aient resté visibles. Ensuite, l'article 2 utilise un espace de recherche réduit dans lequel seuls les trois principaux attributs sont échantillonnés dans le temps, alors que l'article 3 échantillonne tout l'espace 3D photométrique × temps. Comme le début du stimulus de chaque essai dans l'article 2 commence abruptement par un masque de visage complet, que les apparitions des attributs constituent de grosses parties du visage et que le plan de l'image n'est pas complètement échantillonné, il est possible que les sujets aient employé des stratégies atypiques (voir

Jalali et al., 2018). Finalement, la troisième différence importante concerne l'ajustement du nombre de bulles. Dans l'article 2, nous avons opté pour fixer la probabilité que chaque élément soit une bulle à une constante (et donc le nombre de bulles était approximativement constant d'un essai à l'autre et oscillait toujours autour de cette même valeur) afin que les essais soient plus comparables l'un à l'autre. Une approche similaire a été employée dans plusieurs études antérieures utilisant les bulles en neuroimagerie (e.g., Ince et al., 2016; Schyns, Petro & Smith, 2007; Zhan et al., 2019). Notons cependant que ces études ont maintenu le *nombre* de bulles parfaitement constant, ce qui peut avoir la conséquence fâcheuse de créer des dépendances entre différentes parties des masques de bulles[8]. Même l'emploi d'une *probabilité* constante tel que nous l'avons fait dans l'article 2 présente de potentiels problèmes : si les essais sont (en moyenne) constants en termes de la quantité d'information, ils deviennent différents en termes de la performance du sujet (qui deviendra de plus en plus élevée au fil des essais suite à l'apprentissage). Il y a donc là un choix arbitraire à faire : quoi privilégier entre le maintien de la performance et le maintien de la quantité d'information? Maintenir la quantité d'information constante en moyenne présente selon nous un effet négatif important : au fil de l'apprentissage, il devient de moins en moins nécessaire pour les sujets de porter attention à toute l'information présentée pour effectuer la tâche correctement. Ce désavantage risque d'être particulièrement important dans le cas de bulles temporelles : il se pourrait que les sujets n'utilisent que l'information présentée au début des essais. Si ce n'est pas l'information présentée durant toute la durée des essais qui est utilisée, cela pourrait nous empêcher de découvrir des processus d'accumulation et d'intégration de l'information. Nous avons donc décidé dans l'article 3 d'ajuster la quantité d'information à travers les essais (pour maintenir la performance autour de 75%, ce qui est à mi-chemin entre la chance et une performance parfaite), comme ça a été fait dans la plupart des études utilisant les bulles avec le comportement (e.g., Blais et al., 2009; Caplette et al., 2014, 2016, 2017a; Gosselin & Schyns, 2001; Willenbockel et al., 2010; voir également l'article 1 de cette thèse) mais aussi dans quelques études en neuroimagerie (Rutishauser et al., 2011, 2013). L'utilisation d'une probabilité constante

[8] En effet, si le nombre de bulles est constant et que davantage de bulles sont présentées dans un essai dans une région donnée, cela signifie que moins de bulles sont présentées dans cet essai dans les autres régions de l'image. Une plus grande utilisation d'une région donnée pourrait ainsi occasionner une moins grande utilisation — purement artéfactuelle — d'information dans d'autres régions d'une image de classification.

dans l'article 2 pourrait être à l'origine de l'absence d'accumulation visible dans les résultats EEG (voir la section suivante).

5.3 Accumulation et intégration d'information dans le cerveau

Notre paradigme permet théoriquement de visualiser directement l'intégration (i.e. le traitement au même moment) d'information reçue à différents moments durant la fixation[9], et l'accumulation d'information la précédant (l'information reçue plus tôt devrait être accumulée plus longtemps pour être traitée au même moment. Cela nous permettrait également de révéler l'intégration d'information reçue à différents moments (voir Wyart, de Gardelle, Scholl & Summerfield, 2012b, pour une évidence de fluctuations rythmiques pendant l'accumulation d'évidence). Notez que si l'on a une résolution temporelle suffisante au niveau du stimulus, on devrait également pouvoir visualiser une éventuelle intégration partielle d'information, i.e. le traitement au même moment de l'information reçue à différents moments pendant seulement une partie de la fixation, dans certaines régions cérébrales intermédiaires.

Dans l'article 2, nous présentons un résultat qui laisse croire à une possibilité d'accumulation, ou du moins de compression temporelle, dans le cerveau : la dernière composante significative avait lieu de moins en moins de temps après la présentation de l'attribut, selon le moment auquel cet attribut avait été présenté, comme s'il y avait une plus longue période d'accumulation avant ce traitement pour les attributs présentés plus tôt que pour les attributs présentés plus tard. Évidemment, les attributs présentés aux différents moments n'étaient pas non plus traités au *même* moment — ce pour quoi on parle de compression temporelle (partielle), plutôt que d'une intégration.

Nous avons tenté une analyse différente sur les données MEG (Article 3) et nous avons obtenu des résultats prometteurs (une analyse similaire a également été tentée en EEG mais n'a pas donné de résultats, potentiellement suite au moyennage de l'activité sur le scalp; voir section 5.2.4). Nous allons maintenant décrire cette analyse et ces résultats

[9] Nous supposons ici qu'un traitement, au même moment et par la même région cérébrale, d'information reçue à différents moments implique une intégration de cette information. Ce postulat est délibérément libéral et n'implique pas que l'intégration soit nécessairement un processus conjonctif (opérateur logique « et »).

en détails. Nous avons d'abord effectué les régressions régularisées à nouveau, en alignant les données sur la réponse plutôt que sur le début des stimuli. Nous avons ensuite moyenné les coefficients de régression à travers six régions d'intérêt anatomiques définies d'après l'atlas Mindboggle (Klein et al., 2017). Ces régions étaient, pour chaque hémisphère, les suivantes : (1) région occipitale (cortex lingual, cuneus, cortex péri-calcarin et cortex latéral occipital); (2) région pariétale postérieure (cortex pariétal inférieur, cortex pariétal supérieur); et (3) région inférotemporale (cortex inférotemporal, gyrus fusiforme). Deux modèles de surface ont ensuite été définis avec les inégalités suivantes : $a_1 + st < T < a_2 + bt$, où a_1 est l'intercept de la droite définissant le début de la surface, s est la pente de cette droite (fixée soit à 0 ou 1 selon le modèle), t est le moment de présentation, T est le temps dans la MEG depuis le début du stimulus, a_2 est l'intercept de la droite définissant la fin de la surface, et b est la pente de cette droite; a_1, a_2 et b étaient des paramètres libres. Le modèle a été ajusté aux données à l'aide d'un algorithme génétique, avec une population de 300. Cet algorithme essaie de trouver le minimum global d'une fonction étant donné un espace fini de paramètres et ne dépend pas d'un point de départ arbitrairement défini. Dans notre cas, l'algorithme convergeait toujours sur approximativement les mêmes valeurs de paramètres lorsqu'il était ajusté aux mêmes données. Le paramètre a_1 était contraint entre –500 et –100 ms, le paramètre a_2 était contraint entre –450 ms et –50 ms, et le paramètre b était contraint entre –3 et 3; de plus, le modèle a été ajusté aux données avec une contrainte d'inégalité supplémentaire spécifiant que a_1 devait être plus petit que a_2. Pour chaque ROI, attribut facial, tâche et sujet, nous avons calculé un R^2 pour chaque modèle ($s = 0$ ou $s = 1$) et avons soustrait les deux R^2. Nous avons répété les mêmes analyses avec les cartes temps × temps de la distribution de permutations. À la fois les R^2 observés et les R^2 de la distribution de permutations ont été z-scorés à l'aide des R^2 de la distribution de permutations et nous avons établi un seuil statistique Z ($p < .05$, bilatéral, corrigé pour les comparaisons multiples) en utilisant la méthode de la statistique maximale (Holmes et al., 1996). Si la latence de traitement d'une région donnée est linéaire, i.e. si le traitement survient à une latence fixe après la présentation de l'attribut, le modèle à la pente de 1 ($s = 1$) devrait mieux s'ajuster aux données; si le traitement survient à une latence constante, i.e. si le traitement survient au même moment peu importe quand l'information est reçue sur la rétine, le modèle à la pente nulle ($s = 0$) devrait mieux s'ajuster. Si seulement

l'information reçue à un moment (ou approximativement) est traitée, aucun des deux modèles ne devrait mieux s'ajuster que l'autre. Nous avons trouvé que le meilleur modèle à la latence constante s'ajustait mieux aux données que le meilleur modèle à la latence linéaire dans le cortex pariétal postérieur du sujet 1, pour la bouche dans la tâche de reconnaissance d'expression ($R^2 = .31$ vs $.17$; $Z = 4.85$, $p < .05$, bilatéral, corrigé pour les comparaisons multiples; Figure 5.3). Aucune autre différence significative n'a été observée. Ce résultat indique que l'information reçue sur la rétine à plusieurs moments, voire tous les moments, pendant la fixation est traitée au même moment, ou intégrée, dans le cortex pariétal postérieur, du moins chez ce sujet. Le cortex pariétal postérieur (dont le sulcus intrapariétal, LIP, est une partie) est une région où l'accumulation d'information pour prendre une décision a été observée à plusieurs reprises (Hanks et al., 2015; Huk & Shadlen, 2005; Huk, Katz & Yates, 2017). Nous planifions effectuer dans le futur une analyse plus fine, afin de révéler, on l'espère, de l'accumulation chez d'autres sujets. Pour ce faire, nous allons effectuer une procédure de validation croisée dans laquelle nous allons optimiser les sources sélectionnées pour la région d'intérêt.

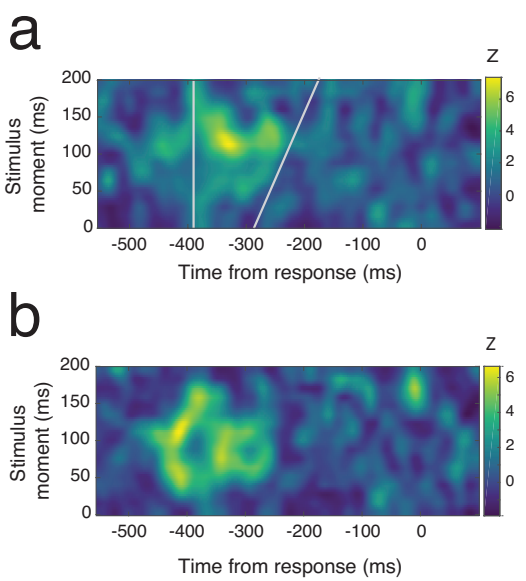


Figure 5.3. A) Illustration de la carte temps × temps pour la bouche dans la tâche d'expression pour la ROI pariétale postérieure du sujet 1. Les lignes grises indiquent le meilleur modèle à pente nulle. **B)** Exemple d'une source de la ROI où l'intégration est particulièrement visible.

5.4 L'effet d'attentes préalables sur l'utilisation d'information

Dans l'article 1 de cette thèse, nous avons évalué comment l'attente d'un objet spécifique modulait notre utilisation de l'information visuelle à travers le temps. Nous avons d'abord observé que l'utilisation des basses FS au début de la fixation était augmentée lorsque les sujets s'attendaient à un objet spécifique; en d'autres mots, que les sujets semblaient commencer à utiliser les basses FS plus tôt lorsqu'il y avait une attente. Comme

l'information disponible est la même dans les deux conditions, l'augmentation de l'utilisation de ces FS s'explique par une différence dans la représentation : maintenant que les sujets s'attendent à un objet spécifique, les attributs (en basses FS) de ces objets sont davantage représentés par les sujets, ce qui explique que la présentation de ces FS corrèle davantage à des réponses correctes dans cette condition. Même si les basses FS sont utilisées tout au long de la fixation dans les deux conditions, les attentes semblent augmenter leur utilisation seulement au début de celle-ci. Plus tard dans la fixation (à la mi-temps), nous avons observé que les attentes augmentaient la variance inter-objets dans l'utilisation des basses FS. Cela signifie que selon l'attente, l'utilisation de ces FS était variable : on a donc ici une démonstration directe que les attentes influencent l'information utilisée de manière spécifique à l'objet/attente. Le fait que l'utilisation soit variable (et davantage variable que s'il n'y a pas d'attente) nous indique que les basses FS ne sont pas utilisées de manière équivalente selon l'objet attendu. Ceci suggère que différents objets sont représentés en FS différentes et que la représentation des basses FS en question est particulièrement variable. Finalement, nous avons corrélé l'utilisation d'information à l'habileté des sujets dans chaque condition. Nous avons observé que les hautes FS à la toute fin de la présentation étaient davantage utilisées par les meilleurs sujets lorsqu'il n'y avait pas d'attente. Les basses FS à la fin étaient davantage utilisées par les meilleurs sujets lorsqu'il y avait une attente. En faisant la différence entre les deux conditions, nous observons que les hautes FS à la toute fin sont moins utilisées par les meilleurs sujets lorsqu'il y a une attente vs lorsqu'il n'y en a pas. Il semblerait ainsi qu'une attente puisse en quelque sorte « remplacer » ces hautes FS, du moins pour les meilleurs sujets. Tel que nous l'avons écrit plus haut, nous avons observé une plus grande utilisation des basses FS en l'absence d'attente. Cela implique que soit (1) les prédictions du modèle de Bar sont fausses et les basses FS ne sont pas utilisées pour créer des hypothèses durant la reconnaissance contrairement à ce qui est postulé dans le modèle de Bar (2003) ou (2) une attente préalable telle qu'utilisée dans notre étude n'est pas équivalente à une hypothèse créée durant la reconnaissance.

Les attentes ont eu des effets distincts à trois moments différents : une augmentation globale de l'utilisation des basses FS au début de la fixation, une variation dans l'utilisation des basses FS selon l'objet attendu à la mi-fixation et une utilisation moins grande des

hautes FS pour les meilleurs sujets à la fin de la fixation. Cette variation temporelle indique une variation dans l'échantillonnage d'information à travers le temps : lorsqu'ils s'attendent à un objet spécifique, les sujets échantillonnent d'abord des attributs (au moins partiellement spécifiques à cet objet) en basses FS, puis des attributs dont les basses FS sont variables selon l'objet. Ce patron d'utilisation reflète une importance particulière des basses FS documentée dans plusieurs études antérieures (Caplette et al., 2016, 2017a; Parker et al., 1992, 1997; Schyns & Oliva, 1994) et une variabilité temporelle de l'utilisation d'information observée à maintes reprises (e.g., Ballard, 2015; Blais et al., 2009, 2012; Cavanagh, 2004; Vinette et al., 2004; Ullman, 1984). Ce décours temporel devrait être reproduit d'une certaine manière dans le cerveau : il serait intéressant d'étudier comment l'information reçue au début et à la fin d'une fixation est traitée par le cerveau lorsqu'une attente est présente (voir section 5.6.4).

5.5 Critiques et limites méthodologiques

Tout travail scientifique présente certaines limites et faiblesses. Nous discutons plusieurs de celles-ci dans la section suivante.

5.5.1 *Bubbles*

Plusieurs critiques peuvent (et ont été) adressées à la technique *Bubbles* dans son ensemble et aux postulats méthodologiques et épistémologiques sur lesquels elle se base (voir également Dupuis-Roy, 2014; Gosselin & Schyns, 2004; Murray & Gold, 2004a, 2004b). La critique la plus commune et la plus sérieuse, à laquelle nous nous attarderons ici, est probablement celle de la vraisemblance des stimuli utilisés. En utilisant du bruit multiplicatif et en échantillonnant seulement une partie des visages, ne présente-t-on pas des stimuli qui n'apparaissent en réalité jamais dans la vie de tous les jours (Murray & Gold, 2004a, 2004b)? Cela est particulièrement flagrant dans le cas de *Bubbles* spatiale où les visages sont visiblement « troués ». Cette critique pourrait également être appliquée dans une moindre mesure à la corrélation inverse classique (et à plusieurs autres techniques

dans une plus ou moins grande mesure); cependant, la corrélation inverse classique se fonde sur l'argument selon lequel le bruit additif est omniprésent dans le système nerveux et qu'ainsi, les transformations appliquées aux stimuli sont de toute manière effectuées par les neurones (Murray & Gold, 2004a). Il s'avère toutefois que le bruit multiplicatif est également présent dans le système nerveux (Lu & Dosher, 1998, 1999). De plus, les objets visuels sont rarement perçus dans leur entièreté dans la vie de tous les jours : ils sont le plus souvent partiellement cachés par d'autres objets se trouvant plus près de nous (Gosselin & Schyns, 2004; Murray, Sekuler & Bennett, 2001). L'échantillonnage aléatoire utilisé par la méthode fait également en sorte qu'utiliser une stratégie atypique (e.g. varier l'utilisation d'information selon l'information présentée) serait sous-optimal. Finalement, les résultats obtenus avec *Bubbles* ont été reproduits à l'aide d'autres techniques expérimentales (e.g., Collin & McMullen, 2005; Gaspar, Sekuler & Bennett, 2008; Goffaux, van Zon & Schiltz, 2011; Nestor & Tarr, 2008a, 2008b; Solomon & Pelli, 1994; Thurman & Grossman, 2011), suggérant que des stratégies atypiques ne sont pas induites.

5.5.1.1 Bubbles *dans la dimension temporelle*

L'utilisation de *Bubbles* dans le domaine temporel spécifiquement peut également être critiquée pour son potentiel manque de validité écologique. En effet, l'input sensoriel n'est pas bloqué de manière intermittente dans la vie de tous les jours tel qu'il l'est avec des bulles temporelles. Des contre-arguments similaires à ceux discutés plus haut peuvent être amenés en réponse à cette critique. D'abord, même si l'input sensoriel ne fluctue pas à travers le temps, son traitement semble fluctuer à travers le temps (Fiebelkorn et al., 2013; Landau & Fries, 2012; VanRullen et al., 2007; voir article 3 et section 5.2.2). Ensuite, l'échantillonnage (temporel) aléatoire prévient en bonne partie l'adoption d'une stratégie atypique selon l'information présentée, car une telle stratégie serait sous-optimale. En effet, non seulement tenter d'utiliser toute information qui apparaît ne mènera pas à de bonnes réponses en moyenne, mais une telle tentative est futile puisqu'une bulle spatiotemporelle ne dure pas suffisamment longtemps pour que l'observateur fasse une saccade oculaire ou attentionnelle vers sa position (en supposant un lissage temporel raisonnable). Cela dit, il est tout de même possible qu'une stratégie atypique soit utilisée et c'est probablement avec les bulles spatiotemporelles qu'une telle critique a le plus de poids. En effet, il est possible

que l'apparition d'une bulle attire l'attention de manière exogène à cet endroit (Jalali et al., 2018). Notons que l'on peut minimiser cette potentielle limite en utilisant des bulles plus petites spatialement et plus grandes et lisses temporellement. De plus, on peut démontrer empiriquement que l'attention n'est pas attirée à chaque fois qu'une bulle est présentée, du moins dans l'article 3 : en effet, nous obtiendrions dans cette situation une activité significative pour l'information reçue à n'importe quel moment, ce qui n'est pas le cas. Cela pourrait cependant potentiellement expliquer certains résultats en EEG (article 2) où les attributs spatiaux échantillonnés sont toujours les mêmes et possèdent une plus grande taille. Nous avons également observé que l'activité est parfois plus grande pour l'information reçue au tout début de la fixation. Il est possible que le début brusque des stimuli provoque cette activité plus grande. Cependant, un tel début brusque est propre à presque toutes les expériences visuelles. Il serait intéressant d'évaluer le traitement visuel dans un contexte où un tel événement est éliminé (voir Lalor et al., 2006; VanRullen & MacDonald, 2012). Notons que la critique de la potentielle utilisation de stratégies atypiques est particulièrement importante dans le cas où les bulles spatiotemporelles sont employées avec la neuroimagerie : en effet, même si la stratégie n'est globalement pas affectée, il est certain que le traitement dans le cerveau est affecté *à quelque part*, puisque l'information sensorielle est différente. Encore une fois cependant, ce fait est vrai pour toute expérience où l'input sensoriel est manipulé (e.g., Fiebelkorn et al., 2018; Helfrich et al., 2018; et si on considère le fait qu'une présentation abrupte d'un stimulus d'une durée prédéterminée est une manipulation, presque toute expérience est concernée). De plus, dans un cas où l'échantillonnage est aléatoire et une telle stratégie est sous-optimale, cet effet risque d'être minimal et limité à des aires de bas niveau et aux aires traitant le mouvement. Finalement, on peut minimiser les biais dus à la méthode en comparant l'information reçue à différents moments (plutôt qu'en regardant uniquement le traitement brut des informations reçues à chaque moment).

Une autre potentielle limite à la technique de *Bubbles* temporelle est le lissage Gaussien : en effet, ce lissage limite notre résolution temporelle et fréquentielle. Cette limite est également présente pour d'autres dimensions (e.g., spatiales). Cependant, comme discuté plus haut, il peut s'agir là d'une force de la méthode. D'abord, ce lissage nous permet d'augmenter notre puissance statistique et d'utiliser des tests statistiques basés sur

la *Random Field Theory* (Chauvin et al., 2005). Ensuite, pour la dimension temporelle spécifiquement, un tel lissage peut minimiser la possibilité d'artéfacts attentionnels puisque l'information apparaît de manière moins abrupte (avec un taux de rafraichissement usuel). Ainsi, il s'agit d'un compromis à faire : réduire la possibilité d'artéfacts attentionnels vs augmenter la résolution temporelle.

Une dernière limite potentielle concerne le caractère arbitraire de la durée de la présentation et des latences associées. En effet, les valeurs absolues des latences (e.g., l'information reçue à 50 ms) obtenues avec cette méthode ont peu de signification et dépendent de la durée de présentation choisie. (Notons cependant, encore une fois, que cette limite est inhérente à toute expérience utilisant des stimuli visuels d'une durée prédéfinie.) Qui plus est, les valeurs relatives des latences (e.g., l'information reçue 50 ms plus tard) pourraient également avoir peu de signification dans certains cas. En effet, si des variations non-oscillatoires sont présentes dans le traitement de l'information (par exemple, un effet de récence ou de primauté, ou une accumulation linéaire d'information jusqu'à l'atteinte d'un seuil prédéterminé), les latences en jeu (e.g., une accumulation pendant 300 ms) seront probablement différentes de ce qui se passe réellement dans le cerveau humain avec un stimulus non-échantillonné : le fait que seulement une partie de l'information est révélée à chaque moment implique nécessairement que l'échantillonnage et l'accumulation d'information prendront plus de temps dans ce cas. En d'autres mots, l'échantillonnage est possiblement étiré temporellement dans une certaine mesure (une raison de plus pour privilégier un nombre de bulles ajusté essai par essai est de maintenir cet « étirement » constant). En réponse à cette critique, notons que cet échantillonnage risque de s'apparenter à une situation réelle où le stimulus est bruité ou faible. De plus, un remède simple à cette potentielle limite est de toujours considérer le temps relativement à la fixation (e.g., début vs fin de la fixation, plutôt que 0 et 200 ms).

5.5.2 Vérification vs catégorisation

Alors que dans les articles 2 et 3, nous avons employé une tâche de catégorisation où chaque alternative correspond à une touche du clavier, nous avons employé dans l'article 1 une tâche de vérification où un nom est présenté en plus de l'objet et où le sujet doit

indiquer si les deux correspondent. Une telle tâche est nécessaire dans un cas où un grand nombre d'alternatives est offert (à moins de laisser les sujets nommer l'objet qu'ils reconnaissent, mais cela comporte aussi son lot de problèmes). Nous avons choisi d'utiliser un grand nombre d'alternatives afin de reproduire plus fidèlement l'identification telle qu'elle s'opère habituellement dans la vie réelle : l'utilisation de seulement quelques alternatives aurait biaisé l'information utilisée par les sujets vers seulement ce qui discrimine les quelques alternatives l'une de l'autre. L'emploi d'une tâche de vérification pour notre étude comporte cependant au moins deux problèmes. D'abord, le traitement de l'information visuelle pourrait être influencé par le nom qui est présenté après l'objet dans la condition *no-expectation*; ainsi, l'utilisation d'information retrouvée par nos analyses ne correspond pas nécessairement parfaitement à ce qui a été uniquement utilisé pendant la fixation même (voir Bang & Rahnev, 2017). Cette limite n'est pas critique pour nos buts, puisque nous comparons deux conditions et obtenons tout de même des différences; cependant, il est possible que nous aurions obtenu davantage de différences si nous avions utilisé une tâche de catégorisation. Ensuite, une telle tâche de vérification est sans doute moins écologique qu'une tâche de catégorisation : nous sommes dans la vie courante plus souvent amenés à identifier un objet (que ce soit implicitement ou explicitement) qu'à vérifier si son identité correspond à une étiquette. Finalement, l'emploi d'une tâche de vérification implique que dans les essais où le stimulus et le nom ne correspondent pas (essais négatifs), le sujet n'est pas obligé de continuer à utiliser l'information dès qu'il a confirmé que l'objet et le stimulus ne correspondent pas, puisqu'il n'a pas à rapporter l'identité du stimulus. Les résultats portant sur les essais négatifs portent donc sur un processus différent de la reconnaissance d'objets classique. Cependant, exclure ces essais est cependant trop coûteux d'un point de vue statistique. Nous planifions reproduire cette étude avec une tâche de catégorisation (voir section 5.6).

5.5.3 Attentes vs attention

Une critique potentielle à l'étude de l'article 1 pourrait être que les différences que nous observons entre les conditions sont dues à l'attention et ne sont donc pas des effets purs des attentes (voir Kok et al., 2012b; Summerfield & Egner, 2009, 2016). Nous ne nous

opposons pas à cette possibilité. Le but de cette étude n'était pas nécessairement d'évaluer uniquement des effets « purs » des attentes (si une telle chose existe), mais bien d'évaluer comment l'attente d'un objet peut influencer l'échantillonnage de différents attributs, que ce soit par l'attention ou par tout autre mécanisme. Nous argumentons qu'il ne s'agit pas là d'une faiblesse et que, du moins à ce niveau, lorsque l'on parle de la modulation de plusieurs attributs différents, les effets des attentes sont indissociables de l'attention. Même lorsqu'il s'agit de la modulation de l'utilisation d'un seul attribut par l'attente de cet attribut (Bang & Rahnev, 2017; Cheadle et al., 2015; Wyart et al., 2012a), la dissociation entre attente et attention est souvent incertaine, malgré les efforts de plusieurs pour orthogonaliser les deux facteurs (e.g., Cheadle et al., 2015; Jiang, Summerfield & Egner, 2013; Kok et al., 2012b; Wyart et al., 2012a; voir également de Lange et al., 2018). En effet, la stratégie employée par ces études consiste à indiquer au début de chaque bloc la probabilité de l'apparition de chaque stimulus à chaque endroit du champ visuel afin de manipuler les attentes et d'indiquer au début de chaque essai l'endroit du champ visuel qui sera pertinent pour la tâche afin de manipuler l'attention. Cependant, rien n'empêche la connaissance des contingences entre les stimuli d'influencer l'attention : en effet, même si on lui indique que la tâche sera probablement à faire sur le stimulus de gauche, rien n'empêche le sujet de porter attention à droite si on lui indique qu'il est plus probable que le stimulus apparaisse à droite, surtout s'il reste une possibilité que la tâche soit à effectuer à droite. Tout cela est sans considérer la sous-optimalité probable des sujets qui pourraient porter davantage attention dans ce cas à droite qu'à gauche, même si ça les menait à une performance plus faible en moyenne. Évidemment, tout cela dépend de notre définition de l'attention. Comme cette définition est assez élusive (e.g., Di Lollo, in press; Tsotsos, 2019), nous préférons nous en tenir à des hypothèses qui ne font pas appel à ce construit.

5.6 Perspectives futures

Les travaux relatés dans cette thèse ouvrent plusieurs nouvelles possibilités, à la fois en termes d'analyses supplémentaires et d'expériences futures. Nous décrivons plusieurs de ces pistes de recherche dans cette section.

5.6.1 Échantillonnage rythmique

Nous planifions effectuer plusieurs analyses supplémentaires afin de caractériser avec plus de précision l'échantillonnage rythmique observé dans l'article 3 et d'approfondir notre compréhension de ce phénomène. D'abord, nous voulons caractériser avec plus de précision ce qui distingue les sources présentant un échantillonnage oscillatoire de celles n'en présentant pas. Des analyses préliminaires révèlent que les oscillations sont surtout présentes sur les sources plus tardives (i.e. commençant le traitement plus longtemps après la présentation de l'attribut). Similairement, nous allons vérifier si cet échantillonnage rythmique varie selon la latence depuis la présentation de l'attribut, sur une même source. Une analyse des phases est également à effectuer : des analyses préliminaires ont révélé qu'il n'y a pas de différence significative de phase entre les différents attributs faciaux (probablement parce qu'il semble plutôt y avoir une différence de fréquence), mais une différence de phase pourrait exister entre différentes régions représentant le même attribut.

Nous souhaitons également tester la présence d'oscillations dans le *traitement* de l'information. Pour ce faire, nous pouvons analyser la dimension du temps dans la MEG dans les cartes temps × temps : cela révélera les oscillations évoquées, c'est-à-dire dont la phase est réinitialisée par la présentation de l'attribut à chaque essai. Nous pouvons également effectuer les analyses de régression sur le signal filtré dans plusieurs bandes de fréquences : des oscillations significatives indiqueront des oscillations dans le traitement d'un attribut dont la phase n'a pas nécessairement été réinitialisée. Il sera intéressant par la suite de vérifier si ces oscillations, s'il y en a, sont dans des fréquences similaires ou distinctes de celles observées dans l'échantillonnage. Il pourrait également être intéressant de vérifier les oscillations présentes dans le signal brut de différentes sources afin de vérifier si celles-ci corrèlent avec l'échantillonnage observé dans les mêmes sources.

Afin de mieux comprendre l'origine et le but de cet échantillonnage rythmique, nous planifions effectuer des analyses de connectivité entre les différentes régions anatomiques et d'effectuer des régressions sur ces patrons de connectivité par essai : nous pourrons ensuite visualiser la connectivité entre deux régions pour de l'information visuelle

spécifique dans une carte temps × temps comme celles utilisées jusqu'ici. Nous allons également vérifier le lien entre cet échantillonnage cérébral et l'échantillonnage comportemental, potentiellement en utilisant des analyses de variance commune (Hebart et al., 2018; Seibold & McPhee, 1979) ou d'information interactive (Ince, 2017; Ince et al., 2017; McGill, 1954), qui nous permettrons de relier stimulus, activité cérébrale et comportement.

5.6.2 Accumulation et intégration de l'information

Pour l'instant, nous n'avons réussi à révéler le traitement simultané d'information reçue à différents moments que chez un sujet (sur trois). Cependant, un tel traitement doit nécessairement avoir lieu à quelque part dans le cerveau avant la réponse du sujet. Il se peut que nos analyses n'aient pas été optimales étant donné notre ratio signal/bruit relativement faible (voir section 5.3). Nous pourrions augmenter la possibilité de découvrir un tel effet en optimisant davantage notre design expérimental pour la MEG (e.g., de Lange, Rahnev, Donner & Lau, 2013; Donner, Siegel, Fries & Engel, 2009) ou en utilisant des enregistrements unicellulaires ou multicellulaires chez le singe (e.g., Gold & Shadlen, 2000; Hanks, Ditterich & Shadlen, 2006; Katz et al., 2016) ou chez des patients épileptiques implantés. Il serait également utile d'utiliser un design expérimental avec des conditions nécessitant explicitement plus ou moins d'accumulation, par exemple en ayant différentes quantités de bruit pour maintenir la performance à différents niveaux.

5.6.3 Expliquer des différences temporelles par l'échantillonnage ou le traitement

Une autre avenue intéressante à emprunter serait de tenter de démontrer empiriquement que des différences dans le temps de traitement peuvent être causées par des différences dans le moment d'échantillonnage (voir également Schoenfeld et al., 2014). Par exemple, nous pourrions utiliser deux attributs que l'on sait habituellement traités à la même vitesse (e.g., les deux yeux d'un visage) et utiliser deux tâches différentes qui forcent l'échantillonnage de l'un avant l'autre (e.g., indiquer si l'œil droit a cligné, seulement dans le cas où l'œil gauche a d'abord cligné, et vice-versa). Pour chacune de ces tâches, il

pourrait y avoir une condition avec *Bubbles* dans l'espace photométrique (bulles spatiales) et une condition avec *Bubbles* dans l'espace photométrique × temps (bulles spatiotemporelles). L'EEG ou la MEG serait utilisées afin d'analyser les moments de traitement dans des régions spécifiques. Nous démontrions (on l'espère) d'abord un traitement plus tardif d'un œil dans chaque tâche à l'aide des bulles spatiales; ensuite, nous pourrions démontrer que ce traitement plus tardif est dû à un échantillonnage plus tardif à l'aide des bulles spatiotemporelles. Alternativement, nous pourrions utiliser les basses et hautes FS comme attributs puisqu'on les sait traitées à des vitesses différentes (e.g., Derrington & Lennie, 1984; Mazer et al., 2002) et naturellement échantillonnées à des moments différents dans une tâche de reconnaissance d'objets courants (e.g., Caplette et al., 2016) mais pouvant être échantillonnées différemment selon la tâche (Schyns & Oliva, 1999). L'utilisation d'un design similaire à celui mentionné ci-haut pourrait permettre de visualiser quelle portion de la différence dans les moments de traitement des basses et hautes FS est due à une différence dans le moment d'échantillonnage et quelle portion est due à une différence dans la vitesse de traitement (voir section 1.3.2).

5.6.4 Influence des attentes sur les représentations

Comme discuté brièvement à la section 5.5.2, nous planifions faire une seconde expérience évaluant l'influence des attentes sur l'utilisation d'information, cette fois-ci en utilisant une tâche de catégorisation. Une telle expérience permettrait d'éliminer quelques limites relatives à la tâche de vérification et de répliquer notre effet avec une tâche différente. De plus, l'utilisation d'un plus petit nombre d'objets nous permettrait d'analyser séparément les images de classification pour chaque objet, et même pour chaque combinaison d'attentes et d'objets. Nous pourrions de plus corréler les FS utilisées pour chaque objet avec le spectre des images originales d'objets (notre ratio signal/bruit est trop faible pour une telle analyse dans l'étude de l'article 1).

Nous pourrions également utiliser la neuroimagerie afin de caractériser le traitement de différents attributs (potentiellement reçus à différents moments) par différentes régions cérébrales lorsqu'un objet spécifique est attendu.

5.6.5 Modulation descendante pendant la reconnaissance

L'expérience mentionnée à la section précédente nous permettrait d'investiguer simultanément une autre question d'intérêt relative au décours temporel de la reconnaissance d'objets : la modulation descendante de l'échantillonnage pendant la reconnaissance même (voir section 1.3.1.2). En effet, la présence de variance inter-objets dans l'utilisation d'information dans la condition sans attente indiquerait nécessairement une modulation de l'utilisation tardive d'information par l'information échantillonnée tôt. Puisque les spectres sont égalisés en amplitude à travers les objets et que les sujets ne connaissent pas d'avance les objets qui seront présentés, une utilisation variable à travers les objets ne peut qu'être due à une telle modulation. Nous avons observé cet effet dans les données présentées dans l'article 1 (voir les z-scores légèrement plus élevés pour presque toutes les FS à partir de 250 ms dans les deux conditions; quelques moments de présentation sont significatifs lorsqu'on moyenne à travers les FS). Cependant, puisqu'une tâche de vérification a été employée dans cette étude (et dans les études antérieures utilisant un paradigme similaire; Caplette et al., 2016, 2017a), nous ne pouvons éliminer la possibilité que cet effet soit dû à une influence du nom d'objet présenté après le stimulus (même si la latence tardive nous indique que ce n'est probablement pas le cas) — d'où l'intérêt d'effectuer une tâche de catégorisation avec aucun événement intermédiaire entre le stimulus et la réponse.

Références

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, *7*(10), 1057–1058.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68–72.

Adolphs, R., Spezio, M. L., Parlier, M., & Piven, J. (2008). Distinct face-processing strategies in parents of autistic children. *Current Biology*, *18*(14), 1090-1093.

Ahumada Jr, A. J. (1996). Perceptual classification images from Vernier acuity masked by noise. *Perception*, *25*(1), 2.

Allen, E. A., & Freeman, R. D. (2006). Dynamic Spatial Processing Originates in Early Visual Pathways. *Journal of Neuroscience*, *26*(45), 11763–11774.

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459–1469.

Baldauf, D., & Desimone, R. (2014). Neural Mechanisms of Object-Based Attention. *Science*, *344*(6182), 424–427.

Ballard, D. H. (2015). *Brain computation as hierarchical abstraction*. Cambridge, USA: MIT Press.

Bang, J. W., & Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Scientific Reports*, *7*(1), 1–12.

Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, A. T. (1997). Is there a" language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual cognition*, *4*(3), 311–331.

Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science*, *364*:453.

Blais, C., Arguin, M., & Gosselin, F. (2013). Human visual processing oscillates: Evidence from a classification image technique. *Cognition*, *128*(3), 353–362.

Blais, C., Fiset, D., Arguin, M., Jolicœur, P., Bub, D., & Gosselin, F. (2009). Reading between Eye Saccades. *PLoS One*, *4*(7):e6448.

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, *50*(12), 2830–2838.

Bobrow, D. G., & Collins, A. (1975). *Representation and understanding.* Amsterdam: Elsevier.

Brisson, B., & Jolicoeur, P. (2007). The N2pc component and stimulus duration. *Neuroreport*, *18*(11), 1163–1166.

Brown, E., & Perrett, D. I. (1993). What gives a face its gender?. *Perception*, *22*(7), 829–840.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research* Reviews, *36*, 96–107.

Bullier, J., & Nowak, L. G. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology*, *5*(4), 497–503.

Bullock, T. H. (1997). Signals and signs in the nervous system: the dynamic anatomy of electrical activity is probably information-rich. *Proceedings of the National Academy of Sciences*, *94*(1), 1–6.

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37), 16048–16053.

Calderone, D. J., Hoptman, M. J., Martínez, A., Nair-Collins, S., Mauro, C. J., Bar, M., et al. (2013). Contributions of low and high spatial frequency processing to impaired object recognition circuitry in schizophrenia. *Cerebral Cortex*, *23*(8), 1849–1858.

Caplette, L., West, G., Gomot, M., Gosselin, F., & Wicker, B. (2014). Affective and contextual values modulate spatial frequency use in object recognition. *Frontiers in Psychology*, *5*:512.

Caplette, L., Wicker, B., & Gosselin, F. (2016). Atypical Time Course of Object Recognition in Autism Spectrum Disorder. *Scientific Reports*, *6*:35494.

Caplette, L., Wicker, B., Gosselin, F., & West, G. (2017a). Hand position alters vision by modulating the time course of spatial frequency use. *Journal of Experimental Psychology: General*, *146*(7), 917–923.

Caplette, L., McCabe, E., Blais, C., & Gosselin, F. (2017b). The Time Course of Object, Scene and Face Categorization. In C. Lefebvre and H. Cohen (Eds.), *Handbook of Categorization in Cognitive Science* (2nd Edition). Amsterdam: Elsevier.

Cavanagh, P. (2004). Attention routines and the architecture or selection. In M. I. Posner (Ed.) *Cognitive Neuroscience of Attention* (pp. 13–28). New York, USA: Gilford Press.

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: sprites. *Cognition*, *80*, 47–60.

Chalk, M., Seitz, A. R., & Series, P. (2010). Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision*, *10*(8):2.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*(9), 659–667.

Cheadle, S., Egner, T., Wyart, V., Wu, C., & Summerfield, C. (2015). Feature expectation heightens visual sensitivity during fine orientation discrimination. *Journal of Vision*, *15*(14):14.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Coleman, J. C. (1949). Facial expressions of emotion. *Psychological Monographs: General and Applied*, *63*(1).

Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Perception & Psychophysics*, *67*(2), 354–364.

Cowan, N. (2000). Processing limits of selective attention and working memory: Potential implications for interpreting. *Interpreting*, *5*(2), 117–146.

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*:604.

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779.

de Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *Journal of Neuroscience*, *33*(4), 1400–1410.

Derrington, A. M., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurons in lateral geniculate nucleus of macaque. *The Journal of Physiology*, *357*, 219–240.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Di Lollo, V. (in press). Attention is a sterile concept; iterative reentry is a fertile substitute. *Consciousness and Cognition*.

Di Russo, F., Stella, A., Spitoni, G., Strappini, F., Sdoia, S., Galati, G., et al. (2012). Spatiotemporal brain mapping of spatial attention effects on pattern-reversal ERPs. *Human Brain Mapping*, *33*(6), 1334–1351.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How Does the Brain Solve Visual Object Recognition? *Neuron*, *73*(3), 415–434.

Donner, T. H., Siegel, M., Fries, P., & Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Current Biology*, *19*(18), 1581–1585.

Duncan, J., Gosselin, F., Cobarro, C., Dugas, G., Blais, C., & Fiset, D. (2017). Orientations for the successful categorization of facial expressions and their link with facial features. *Journal of Vision*, *17*(14), 7–16.

Dunlap, K. (1927). The role of eye-muscles and mouth-muscles in the expression of the emotions. *Genetic Psychology Monographs*, *2*(3), 196–233.

Dupuis-Roy, N. (2014). *Information utile à la catégorisation du sexe des visages* (Doctoral dissertation). Retrieved from http://hdl.handle.net/1866/11590

Dupuis-Roy, N., Faghel-Soubeyrand, S., & Gosselin, F. (2019). Time course of the use of chromatic and achromatic facial information for sex categorization. *Vision Research*, *157*, 36–43.

Dupuis-Roy, N., Fortin, I., Fiset, D., & Gosselin, F. (2009). Uncovering gender discrimination cues in a realistic setting. *Journal of Vision*, *9*(2):10.

Estephan, A. E. X., Fiset, D., Saumure, C., Plouffe-Demers, M.-P., Zhang, Y., Sun, D., & Blais, C. (2018). Time Course of Cultural Differences in Spatial Frequency Use for Face Identification. *Scientific Reports*, *8*:1816.

Esterman, M., & Yantis, S. (2010). Perceptual Expectation Evokes Category-Selective Cortical Activity. *Cerebral Cortex*, *20*(5), 1245–1253.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47.

Fiebelkorn, I. C., Pinsk, M. A., & Kastner, S. (2018). A Dynamic Interplay within the Frontoparietal Network Underlies Rhythmic Spatial Attention. *Neuron*, *99*(4), 842–853.

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic Sampling within and between Objects despite Sustained Attention at a Cued Location. *Current Biology*, *23*(24), 2553–2558.

Fiset, D., Blais, C., Arguin, M., Tadros, K., Éthier-Majcher, C., Bub, D., & Gosselin, F. (2009). The spatio-temporal dynamics of visual letter recognition. *Cognitive Neuropsychology*, *26*(1), 23–35.

Fiset, D., Blais, C., Éthier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for Identification of Uppercase and Lowercase Letters. *Psychological Science*, *19*(11), 1161–1168.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 607–608.

Frois-Wittman, J. (1930). The judgment of facial expression. *Journal of Experimental Psychology*, *13*(2), 113.

Gaspar, C., Sekuler, A. B., & Bennett, P. J. (2008). Spatial frequency tuning of upright and inverted face identification. *Vision Research*, *48*(28), 2817–2826.

Gibson, B. M., Lazareva, O. F., Gosselin, F., Schyns, P. G., & Wasserman, E. A. (2007). Nonaccidental Properties Underlie Shape Recognition in Mammalian and Nonmammalian Vision. *Current Biology*, *17*(4), 336–340.

Goffaux, V., van Zon, J., & Schiltz, C. (2011). The horizontal tuning of face perception relies on the processing of intermediate and high spatial frequencies. *Journal of Vision*, *11*(10):1.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*(6776), 390.

Gosselin, F., Bacon, B. A., & Mamassian, P. (2004) Internal surface representations approximated by reverse correlation. *Vision Research*, *44*(21), 2515–2520.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.

Gosselin, F., & Schyns, P. G. (2002). RAP: a new framework for visual categorization. *Trends in Cognitive Sciences*, *6*(2), 70–77.

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science*, *14*(5), 505–509.

Gosselin, F., & Schyns, P. G. (2004). No troubles with bubbles: a reply to Murray and Gold. *Vision Research*, *44*(5), 471–477.

Gosselin, F., & Schyns, P. G. (2005). Bubbles: A user's guide. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building Object Categories in Developmental Time* (pp. 109–124). Boca Raton, USA: CRC Press.

Gregoriou, G. G., Gotts, S. J., Zhou, H., & Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science*, *324*(5931), 1207–1210.

Hanks, T. D., & Summerfield, C. (2017). Perceptual Decision Making in Rodents, Monkeys, and Humans. *Neuron*, *93*(1), 15–31.

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neuroscience*, *9*(5), 682–689.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, *520*(7546), 220–223.

Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: implications for perceptual categorization hierarchies. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(4), 1264–1273.

Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The representational dynamics of task and object processing in humans. *eLife*, *7*:e32816.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665–667.

Holcombe, A. O., & Chen, W. Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to <3 Hz when tracking three. *Journal of Vision*, *13*(1):12.

Hong, H., Yamins, D. L. K., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, *19*(4), 613–622.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, *8*(3), 197–230.

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience*, *25*(45), 10420–10436.

Huk, A. C., Katz, L. N., & Yates, J. L. (2017). The Role of the Lateral Intraparietal Area in (the Study of) Decision Making. *Annual Review of Neuroscience*, *40*(1), 349–372.

Ince, R. A. A. (2017). Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy*, *19*(7):318.

Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, *38*(3), 1541–1573.

Ince, R. A. A., Jaworska, K., Gross, J., Panzeri, S., van Rijsbergen, N. J., Rousselet, G. A., & Schyns, P. G. (2016). The Deceptively Simple N170 Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres. *Cerebral Cortex*, *26*(11), 4123–4135.

Ince, R. A. A., van Rijsbergen, N. J., Thut, G., Rousselet, G. A., Gross, J., Panzeri, S., & Schyns, P. G. (2015). Tracing the Flow of Perceptual Features in an Algorithmic Brain Network. *Scientific Reports*, *5*:17681.

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, *141*(1), 19–25.

Jalali, S., Martin, S. E., Murphy, C. P., Solomon, J. A., & Yarrow, K. (2018). Classification Videos Reveal the Visual Information Driving Complex Real-World Speeded Decisions. *Frontiers in Psychology*, *9*:2229.

James, W. (1890). *The principles of psychology* (Vol. 1, No. 2). London, UK: Macmillan.

Jiang, J., Summerfield, C., & Egner, T. (2013). Attention Sharpens the Distinction between Expected and Unexpected Percepts in the Visual Brain. *Journal of Neuroscience*, *33*(47), 18438–18447.

Jolicoeur, P., Ullman, S., & Mackay, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition*, *14*(2), 129–140.

Jolicoeur, P., Ullman, S., & Mackay, M. (1991). Visual curve tracing properties. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4), 997–1022.

Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. A dynamic causal modeling study. *Brain & Cognition*, *99*, 46–56.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLoS Computational Biology*, *10*(11):e1003915.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., et al. (2017). Mindboggling morphometry of human brains. *PLoS Computational Biology*, *13*(2):e1005350.

Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior Expectations Bias Sensory Representations in Visual Cortex. *Journal of Neuroscience*, *33*(41), 16275–16284.

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554.

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*, *114*(39), 10473–10478.

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2011). Attention Reverses the Effect of Prediction in Silencing Sensory Signals. *Cerebral Cortex*, *22*(9), 2197–2206.

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science*, *1*(1), 417–446.

Lalor, E. C., Pearlmutter, B. A., & Foxe, J. J. (2009). Reverse Correlation and the VESPA Method. In T. C. Handy (Ed.), *Brain signal analysis: advances in neuroelectric and neuromagnetic methods* (pp. 1–19). Cambridge, USA: MIT Press.

Lalor, E. C., Pearlmutter, B. A., Reilly, R. B., McDarby, G., & Foxe, J. J. (2006). The VESPA: A method for the rapid estimation of a visual evoked potential. *NeuroImage*, *32*(4), 1549–1561.

Landau, A. N., & Fries, P. (2012). Attention Samples Stimuli Rhythmically. *Current Biology*, *22*(11), 1000–1004.

Landau, A. N., Schreyer, H. M., Van Pelt, S., & Fries, P. (2015). Distributed attention is implemented through theta-rhythmic gamma modulation. *Current Biology*, *25*(17), 2332-2337.

Langner, O., Becker, E. S., & Rinck, M. (2009). Social Anxiety and Anger Identification: Bubbles Reveal Differential Use of Facial Information With Low Spatial Frequencies. *Psychological Science*, *20*(6), 666–670.

Lehky, S. R., & Tanaka, K. (2016). ScienceDirect Neural representation for object recognition in inferotemporal cortex. *Current Opinion in Neurobiology*, *37*, 23–35.

Leutgeb, S., & Leutgeb, J. K. (2007). Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map. *Learning & Memory*, *14*(11), 745–757.

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision research*, *38*(9), 1183–1198.

Lu, Z. L., & Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A*, *16*(3), 764–778.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, *81*(1), B1–B9.

Mandel, M. I., Yoho, S. E., & Healy, E. W. (2016). Measuring time-frequency importance functions of speech with bubble noise. *The Journal of the Acoustical Society of America*, *140*(4), 2542–2553.

Mangini, M., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science*, *28*(2), 209–226.

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical transactions of the Royal Society of London. B: Biological sciences*, *262*(841), 23–81.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* New York, USA: Henry Holt and Co.

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1645–1650.

McAlonan, K., Cavanaugh, J., & Wurtz, R. H. (2008). Guarding the gateway to cortex with attention in visual thalamus. *Nature*, *456*(7220), 391–394.

McGill, W. (1954). Multivariate information transmission. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 93–111.

Miller, J. (1988). Discrete and continuous models of human information processing: Theoretical distinctions and empirical results. *Acta Psychologica*, *67*, 191–257.

Morin-Duchesne, X., Gosselin, F., Fiset, D., & Dupuis-Roy, N. (2014). Paper features: A neglected source of information for letter recognition. *Journal of Vision*, *14*(13):11.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological cybernetics*, *66*(3), 241–251.

Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, *11*(5), 2–2.

Murray, R. F., & Gold, J. M. (2004a). Troubles with bubbles. *Vision Research*, *44*(5), 461–470.

Murray, R. F., & Gold, J. M. (2004b). Reply to Gosselin and Schyns. *Vision Research*, *44*(5), 479–482.

Murray, R. F., Sekuler, A. B., & Bennett, P. J. (2001). Time course of amodal completion revealed by a shape discrimination task. *Psychonomic Bulletin & Review*, *8*(4), 713–720.

Nagai, M., Bennett, P. J., & Sekuler, A. B. (2007). Spatiotemporal templates for detecting orientation-defined targets. *Journal of Vision*, *7*(8), 11–16.

Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience*, *5*(8), 812–816.

Neri, P., & Levi, D. (2008). Temporal dynamics of directional selectivity in human vision. *Journal of Vision*, *8*(1):22.

Neri, P., & Levi, D. M. (2007). Temporal Dynamics of Figure-Ground Segregation in Human Vision. *Journal of Neurophysiology*, *97*(1), 951–957.

Nestor, A., & Tarr, M. J. (2008a). Gender Recognition of Human Faces Using Color. *Psychological Science*, *19*(12), 1242–1246.

Nestor, A., & Tarr, M. J. (2008b). The segmental structure of faces and its use in gender recognition. *Journal of Vision*, *8*(7):7.

Nowak, L. G., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual neuroscience*, *12*(2), 371–384.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*(11), 1203–1209.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452–454.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*(1), 72–107.

Palmer, S. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorization* (pp. 259–303). Mahwah, USA: Lawrence Erlbaum Associates.

Panzeri, S., Brunel, N., Logothetis, N. K., & Kayser, C. (2010). Sensory neural codes using multiplexed temporal scales. *Trends in Neurosciences*, *33*(3), 111–120.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*(2), 147–160.

Parker, D. M., Lishman, J. R., & Hughes, J. (1996). Role of coarse and fine spatial information in face and object processing. *Journal of Experimental Psychology. Human Perception and Performance*, *22*(6), 1448–1466.

Parker, D. M., Lishman, J. R., & Hughes, J. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception*, *26*(9), 1169–1180.

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8):13.

Puri, A. M., & Wojciulik, E. (2008). Expectation both helps and hinders object perception. *Vision Research*, *48*(4), 589–597.

Purushothaman, G., Chen, X., Yampolsky, D., & Casagrande, V. A. (2014). Neural mechanisms of coarse-to-fine discrimination in the visual cortex. *Journal of Neurophysiology*, *112*(11), 2822–2833.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science*, *28*(2), 147–166.

Romei, V., Driver, J., Schyns, P. G., & Thut, G. (2011). Rhythmic TMS over Parietal Cortex Links Distinct Brain Frequencies to Global versus Local Visual Processing. *Current Biology*, *21*(4), 334–337.

Ruckmick, C. A. (1921). A preliminary study of the emotions. *Psychological monographs*, *30*(3), 30–35.

Rungratsameetaweemana, N., Itthipuripat, S., Salazar, A., & Serences, J. T. (2018). Expectations Do Not Alter Early Sensory Processing during Perceptual Decision-Making. *Journal of Neuroscience*, *38*(24), 5632–5648.

Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception*, *32*(9), 1093–1107.

Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception*, *32*(9), 1093–1107.

Rutishauser, U., Tudusciuc, O., Neumann, D., Mamelak, A. N., Heller, A. C., Ross, I. B., et al. (2011). Single-unit responses selective for whole faces in the human amygdala. *Current Biology*, *21*(19), 1654–1660.

Rutishauser, U., Tudusciuc, O., Wang, S., Mamelak, A. N., Ross, I. B., & Adolphs, R. (2013). Single-Neuron Correlates of Atypical Face Processing in Autism. *Neuron*, *80*(4), 887–899.

Schoenfeld, M. A., Hopf, J.-M., Merkel, C., Heinze, H.-J., & Hillyard, S. A. (2014). Object-based attention involves the sequential activation of feature-specific cortical modules. *Nature Neuroscience*, *17*(4), 619–624.

Schyns, P. G., & Ince, R. A. (submitted). Making the brain-activity-to-information leap using a novel framework: Stimulus Information Representation (SIR).

Schyns, P. G., & Oliva, A. (1994). From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition. *Psychological Science*, *5*(4), 195–200.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*(3), 243–265.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409.

Schyns, P. G., Thut, G., & Gross, J. (2011). Cracking the Code of Oscillatory Activity. *PLOS Biology*, *9*(5):e1001064.

Seibold, D. R., & McPhee, R. D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analyses. *Human Communication Research*, *5*(4), 355–365.

Smith, F. W., Muckli, L., Brennan, D., Pernet, C., Smith, M. L., Belin, P., et al. (2008). Classification images reveal the information sensitivity of brain voxels in fMRI. *NeuroImage*, *40*(4), 1643–1654.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological science*, *16*(3), 184–189.

Smith, M. L., Fries, P., Gosselin, F., Goebel, R., & Schyns, P. G. (2009). Inverse mapping the neuronal substrates of face categorizations. *Cerebral Cortex*, *19*(10), 2428–2438.

Smith, M. L., Gosselin, F., & Schyns, P. G. (2006). Perceptual moments of conscious visual experience inferred from oscillatory brain activity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(14), 5626–5631.

Smith, M. L., Gosselin, F., & Schyns, P. G. (2007). From a face to its category via a few information processing states in the brain. *NeuroImage*, *37*(3), 974–984.

Smith, M. L., Gosselin, F., & Schyns, P. G. (2012). Measuring internal representations from behavioral and brain data. *Current Biology*, *22*(3), 191–196.

Smith, N. J., & Kutas, M. (2014). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, *52*(2), 169–181.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, *369*(6479), 395.

Sotiropoulos, G., Seitz, A. R., & Seriès, P. (2011). Changing expectations about speed alters perceived motion direction. *Current Biology*, *21*(21), R883–R884.

Spezio, M. L., Adolphs, R., Hurley, R. S., & Piven, J. (2007). Analysis of face gaze in autism using "Bubbles". *Neuropsychologia*, *45*(1), 144–151.

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, *144*(6), 1089–1104.

Sterzer, P., Frith, C., & Petrovic, P. (2008). Believing is seeing: expectations alter visual awareness. *Current Biology*, *18*(16), R697–R698.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409.

Summerfield, C., & Egner, T. (2016). Feature-Based Attention and Feature-Based Expectation. *Trends in Cognitive Sciences*, *20*(6), 401–404.

Summerfield, C., & Koechlin, E. (2008). A Neural Representation of Prior Information during Perceptual Inference. *Neuron*, *59*(2), 336–347.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive Codes for Forthcoming Perception in the Frontal Cortex. *Science*, *314*(5803), 1311–1314.

Sun, J., & Perona, P. (1998). Where is the sun?. *Nature neuroscience*, *1*(3), 183–184.

Suzuki, M., & Noguchi, Y. (2013). Reversal of the face-inversion effect in N170 under unconscious visual processing. *Neuropsychologia*, *51*(3), 400–409.

Tadros, K., Dupuis-Roy, N., Fiset, D., Arguin, M., & Gosselin, F. (2013). Reading laterally: the cerebral hemispheric use of spatial frequencies in visual word recognition. *Journal of Vision*, *13*(1), 4.

Tanskanen, T., Näsänen, R., Ojanpää, H., & Hari, R. (2007). Face recognition and cortical responses: effect of stimulus duration. *Neuroimage*, *35*(4), 1636–1644.

Tardif, J., Fiset, D., Zhang, Y., Estéphan, A., Cai, Q., Luo, C., ... & Blais, C. (2017). Culture shapes spatial frequency tuning for face identification. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 294–306.

Thurman, S. M., & Grossman, E. D. (2008). Temporal "Bubbles" reveal key features for point-light biological motion perception. *Journal of Vision*, *8*(3):28.

Thurman, S. M., & Grossman, E. D. (2011). Diagnostic spatial frequencies and human efficiency for discriminating actions. *Attention, Perception, & Psychophysics*, *73*(2), 572–580.

Tsotsos, J. K. (2019). Attention: The Messy Reality. *Yale Journal of Biology and Medicine*, *92*, 127–137.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643.

Ullman, S. (1984). Visual Routines. *Cognition*, *18*, 97–159.

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, *5*(1), 1–11.

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*:365.

VanRullen, R. (2016). Perceptual Cycles. *Trends in Cognitive Sciences*, *20*(10), 723–735.

VanRullen, R., & Macdonald, J. S. P. (2012). Perceptual Echoes at 10 Hz in the Human Brain. *Current Biology*, *22*(11), 995–999.

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19204–19209.

Venezia, J. H., Hickok, G., & Richards, V. M. (2016). Auditory "bubbles": Efficient classification of the spectrotemporal modulations essential for speech intelligibility. *The Journal of the Acoustical Society of America*, *140*(2), 1072–1088.

Verstraten, F. A., Cavanagh, P., & Labianca, A. T. (2000). Limits of attentive tracking reveal temporal properties of attention. *Vision research*, *40*(26), 3651–3664.

Victor, J. D. (2000). How the brain uses time to represent and process visual information. *Brain research*, *886*(1-2), 33–46.

Vinette, C., Gosselin, F., & Schyns, P. (2004). Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. *Cognitive Science*, *28*(2), 289–301.

Wang, S., Tudusciuc, O., Mamelak, A. N., Ross, I. B., Adolphs, R., & Rutishauser, U. (2014). Neurons in the human amygdala selective for perceived emotion. *Proceedings*

of the National Academy of Sciences of the United States of America, *111*(30), E3110–E3119.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A*, *4*(10), 2006–2021.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., et al. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology. Human Perception and Performance*, *36*(1), 122–135.

Wyart, V., Nobre, A. C., & Summerfield, C. (2012a) Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3593–3598.

Wyart, V., de Gardelle, V., Scholl, J., & Summerfield, C. (2012b). Rhythmic Fluctuations in Evidence Accumulation during Decision Making in the Human Brain. *Neuron*, *76*(4), 847–858.

Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8619–8624.

Yates, J. L., Pillow, J. W., Huk, A. C., & Katz, L. N. (2016). Dissociated functional significance of decision- related activity in the primate dorsal stream. *Nature*, *535*(7611), 285–288.

Zhan, J., Ince, R. A. A., van Rijsbergen, N., & Schyns, P. G. (2019). Dynamic Construction of Reduced Representations in the Brain for Perceptual Decision Behavior. *Current Biology*, *29*, 319–326.

Annexe A

Article supplémentaire

Real-world expectations and their affective value modulate object processing

Laurent Caplette[1], Frédéric Gosselin[1], Martial Mermillod[2] et Bruno Wicker[1,3]

[1]Département de psychologie, Université de Montréal, Montréal, Québec, H3C 3J7, Canada
[2]LPNC, CNRS & Université Grenoble Alpes, 38058 Grenoble, France
[3]LNC, CNRS & Aix-Marseille Université, 13331 Marseille, France

Keywords: expectations, predictions, object recognition, fMRI, emotion.

Corresponding author: Laurent Caplette
Email: laurent.caplette@umontreal.ca
Address: Department of Psychology, University of Montreal, C.P. 6128 succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada

Abstract

It is well known that expectations influence how we perceive the world. Yet the neural mechanisms underlying this process remain unclear. Studies have focused so far on artificial contingencies between simple neutral cues and events. Real-world expectations are however often generated from complex associations between potentially affective contexts and objects learned over a lifetime. In this study, we used fMRI to investigate how object processing is influenced by neutral and affective context-based expectations. First, we show that the precuneus, the inferotemporal cortex and the frontal cortex are more active during object recognition when expectations have been elicited a priori, irrespectively of their validity or their affective intensity. This result supports previous hypotheses according to which these brain areas integrate contextual expectations with object sensory information. Notably, these brain areas are different from those responsible for simultaneous context-object interactions, dissociating the two processes. Then, we show that early visual areas, on the contrary, are more active during object recognition when no prior expectation has been elicited by a context. Lastly, BOLD activity was shown to be enhanced in early visual areas when objects are less expected, but only when contexts are neutral; the reverse effect is observed when contexts are affective. This result supports recent proposals that affect modulates predictions in the brain. Together, our results help elucidate the neural mechanisms of real-world expectations.

We expect to find hairdryers in bathrooms, tombstones in cemeteries, and baguettes in bakeries, but more rarely tombstones in bathrooms, refrigerators in cemeteries and hairdryers in bakeries. That is, we live in a world where most objects are associated with specific contexts. Throughout a lifetime of experiences, we come to learn these associations, which lead us to form expectations about the objects to be encountered when we navigate the world.

Perception can be understood as the process of integrating such top-down expectations with incoming sensory information. It has been proposed that predictions from high-level areas are transmitted to adjacent lower-level areas and compared with incoming sensory signals, such that only the discrepancy between these two signals – the prediction error – is transmitted up the visual hierarchy (Friston, 2005; see also Mumford, 1992; Ullman, 1995; Rao and Ballard, 1999). In support of this model, expectation of a visual stimulus elicits a specific pattern of activity in the primary visual cortex (Kok, Failing & de Lange, 2014; Kok, Mostert & de Lange, 2017; Hindy, Ng & Turk-Browne, 2016) and the perception of an expected stimulus results in reduced neural activity in sensory cortices (Summerfield, Trittschuh, Monti, Mesulam & Egner, 2008; den Ouden, Daunizeau, Roiser, Friston & Stephan, 2010; Alink, Schwiedrzik, Kohler, Singer & Muckli, 2010; Kok, Jehee & de Lange, 2012a; Todorovic & de Lange, 2012; see de Lange, Heilbron & Kok, 2018, for a review). Some predictions, however, may require a different mechanism than feedback from adjacent visual areas (Hindy et al., 2016): for instance, the hippocampus has been shown to play a role in the generation of predictions (Hindy et al., 2016; Kok & Turk-Browne, 2018), and there is some evidence that parahippocampal (PHC) and retrosplenial (RSC) cortices initiate context-based expectations (Bar, 2003; Bar & Aminoff, 2003; Bar, 2004; Bar et al., 2006; Livne & Bar, 2016; Brandman & Peelen, 2017).

Most studies examining the effect of predictions on perception have used very simple cues such as tones (Summerfield & Koechlin, 2008; den Ouden et al., 2010; Kok et al., 2012a, 2017) or a repetition of the same object (Summerfield et al., 2008; Todorovic & de Lange, 2012). By contrast, expectations about everyday objects usually stem from the surrounding context. Several previous studies investigated context-object relationships, but they used a simultaneous presentation of the object and the scene (Goh et al., 2004;

Jenkins, Yang, Goh, Hong & Park, 2010; Kirk, 2008; Rémy, Vayssière, Pins, Boucart & Fabre-Thorpe, 2014), which makes it hard to disentangle scene-object interactions from scene-based predictions (which occur prior to the object's recognition). To our knowledge, this is the first study to explore sequential context-object interactions.

Relatedly, the effect of predictions has not been considered in the setting of an ecological object recognition task. Simple detection tasks (Jiang, Summerfield & Egner, 2013), delayed discrimination tasks (Kok et al., 2012a, 2014, 2017) or categorization tasks using few alternatives (den Ouden et al., 2010; Kok, Rahnev, Jehee, Lau & de Lange, 2012b) are typically used. Moreover, previous studies on prediction have manipulated predictability by artificial means, either by repeating and alternating stimuli (Summerfield et al., 2008; Todorovic & de Lange, 2012), by having stimuli appearing after different cues with different probabilities during the experiment (den Ouden et al., 2010; Kok et al., 2012a, 2012b, 2014, 2017; Jiang et al., 2013), or by developing arbitrary contingencies shortly before the experiment (Hindy et al., 2016). Associations between contexts and objects formed over a lifetime of experiences may involve mechanisms distinct from these. For instance, real-world expectations are often tinted by some affective value. A visual context can elicit emotional reactions that may influence the recognition of objects in the scene (Lebrecht, Barr, Barrett & Tarr, 2012). In an emotional context (e.g., a cemetery), the affective value may be partially processed before the scene's objects (e.g., a tombstone) and contribute to the object's recognition (Barrett & Bar, 2009). Alternatively, the prediction's affective value might interact with its validity: this is likely to result in a reversal of the prediction error effect in the brain (Miller & Clark, 2018).

In the present study, we aimed to address these shortcomings by investigating how realistic object recognition mechanisms are influenced by task-irrelevant high-level expectations generated by a predictive or non-predictive visual context. The use of everyday objects and scenes allowed us to use associations between objects and contexts formed over a lifetime of experiences, and to compare affective and neutral expectations.

Materials and Methods

Participants

Seventeen healthy adults (9 female; mean age = 24.8; SD = 4.3) were recruited on the campus of Aix-Marseille Université. Participants did not suffer from any neurological, psychological or psychiatric disorder and were free of medication. The experimental protocol was approved by the ethics board of CPP Sud-Méditerranée 1 and the study was carried in accordance with the approved guidelines. Written informed consent was obtained from all participants after the procedure had been fully explained, and a monetary compensation was provided upon completion of the experiment.

Stimuli

In a first validation study, 35 different subjects were shown thirty-three context names and had to give the names of three objects with a high probability of being present in that context. Then, thirty-three public domain scene color images were selected from the internet as context images (see examples in Figure 1a). Context images were selected to ensure that their three most associated objects did not appear in them while still being representative of the context category. In a second validation study, an independent sample of 22 subjects identified what they thought the context images represented (to confirm that the image represented the context), indicated if the context elicited an emotion and, if so, what were its valence (negative to positive, from 0 to 10) and intensity (no emotion to very intense emotion, from 0 to 10). Following this study, 32 visual scenes were selected (one scene was excluded) and split in Affective (e.g., cemetery, beach, luxury hotel) and Neutral (e.g., swimming pool, airport, kitchen) categories at the median of the intensity scores (5.19); valence was not included in the experimental design. On average, neutral contexts had an intensity of 3.19 and a valence of 6.14; affective contexts had an intensity of 6.10 and a valence of 5.53.

Ninety-six color images of objects corresponding to the three most cited names for each context were then selected for the experiment (e.g., swimsuit, diving board and pool ladder for swimming pool; Figure 1a, Table S1). For each context, the experimenters also chose three non-associated objects (selected from the ones that had never been associated

with the context in the second validation study). Every object was the associated object of only one context and the non-associated object of only one other context; moreover, for each context, each one of the three non-associated objects was associated with a different context. A third and final validation study was conducted to collect quantitative measures of the associations between objects and contexts. Forty-four new subjects indicated on a scale from 0 to 10 how much each object was associated to its predictive and non-predictive contexts (context-object pairs were randomized). Measures were z-scored within each subject and averaged across them.

Finally, we randomized the phases of the mean of the context images in the Fourier domain – separately for each RGB color channel – to obtain 96 different phase-scrambled images.

Data acquisition

Functional imaging data were acquired with an ADVANCE 3 Tesla scanner (Bruker Inc., Ettlingen, Germany) equipped with a 2-channel head-coil. Functional images sensitive to BOLD contrast were acquired with a T2*-weighted gradient echo EPI sequence (TR 2400 ms, TE 30 ms, matrix 64 x 64 mm, FOV 192 mm, flip angle 81.6°). Thirty-six slices with a slice gap of 0 mm were acquired within the TR; voxels were 3 x 3 x 3 mm. Between 303 and 311 volumes were acquired in each run, excluding the six dummy scans acquired at the beginning of each run for signal stabilization. Additionally, a high resolution (1 x 1 x 1 mm) structural scan was acquired from each participant with a T1-weighted MPRAGE sequence.

Experimental Design

The LabVIEW (National Instruments Inc., Austin, TX, USA) software was used to project stimuli during the experiment. Stimuli were projected to a screen positioned in the back of the scanner using a video projector. Subjects could see the video reflected in a mirror (15 x 9 cm) suspended 10 cm in front of their face and subtending visual angles of 42 degrees horizontally and 32 degrees vertically.

Each trial was built as follows: a large cue image (see below) spanning the whole screen during 1 s, a black screen during 1.5 to 4 s (duration randomly selected from a

truncated exponential distribution with mean of 2 s), a centered object image on a black background during 133 ms, a black screen during 1.5 to 4 s, and an object name on a black background shown until the subject answered or for a maximum of 1 s (Figure 1b). Subjects answered by pressing one of two buttons on a hand-held response device to indicate if the name corresponded to the object, which occurred on 80% of the trials. A black screen was displayed for an additional 1 s between trials.

On a third of the trials (Predictive condition), the cue image was a scene associated with the object following it (e.g., an airport and a suitcase); on another third (Non-Predictive condition), it was a scene not associated with the object following it (e.g., a church and a tennis racket); on the final third (No-Context condition), it was a scrambled image (always a different one). Each object was shown once in each of these conditions, for a total of 288 trials. Furthermore, Predictive and Non-Predictive conditions were each split evenly into Affective and Neutral subconditions, following the affective intensity of the context. There was therefore a total of 5 conditions: Predictive Affective (or Pred-Aff for short), Predictive Neutral (Pred-Neut), Non-Predictive Affective (noPred-Aff), Non-Predictive Neutral (noPred-Neut) and No-Context (noCont).

The order of trials was randomized. Randomized trials were divided in 3 fixed functional data acquisition runs of 96 trials. Each functional run lasted between 10 and 12 mins, with short breaks between them. The order of the 3 runs was counterbalanced across subjects.

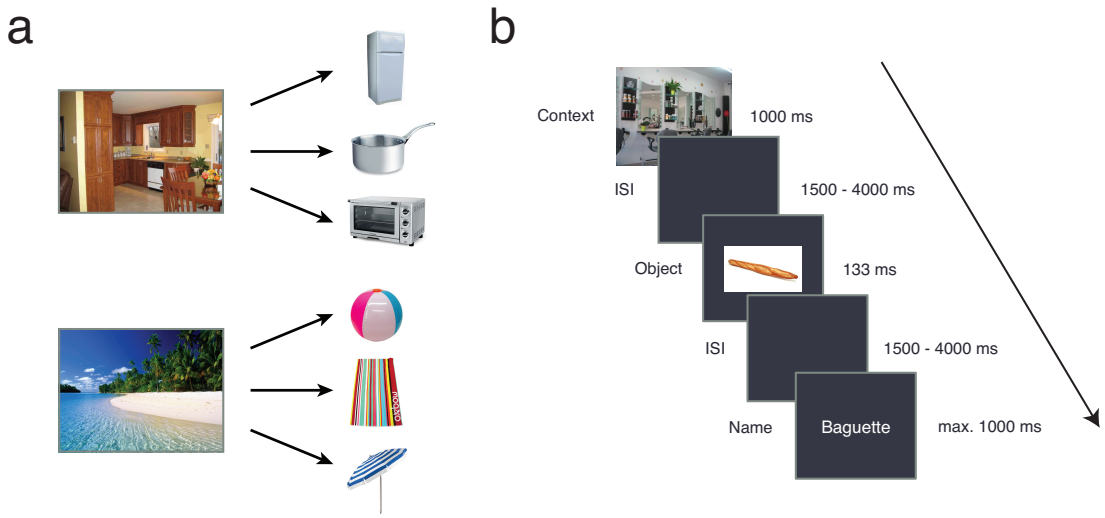


Figure 1. A) Example of one neutral and one affective scene, with their associated objects. See table S1 for a list of all contexts and associated objects. **B)** Example of a trial (Non-Predictive Neutral condition). The object image and name have been enlarged for better viewing.

Data Preprocessing and Analysis

The SPM8 software (http://www.fil.ion.ucl.ac.uk/spm/), running in the MATLAB environment (Mathworks Inc., Natick, MA, USA), was used. T1-weighted structural images were segmented into white matter, gray matter and cerebrospinal fluid, and warped into MNI space. Functional images were realigned, unwarped and corrected for geometric distortions using the field map of each participant, slice time corrected, coregistered to the structural image of the corresponding participant, and smoothed using a 6 mm FWHM isotropic Gaussian kernel.

A standard GLM analysis was performed for each subject. Three events were modelled on each trial: contexts (or scrambled images), objects and object names. Object events (the regressors of interest) were modelled for each condition separately (Pred-Aff, Pred-Neut, noPred-Aff, noPred-Neut and noCont); scene events (regressors of no interest) were also modelled separately for each condition; one additional regressor was included for the object names. All these events were modelled as Dirac delta functions (duration of zero) convolved with SPM8's canonical hemodynamic response function. To get rid of potential effects caused by differences in context-object associations, we included an

additional parametric regressor which consisted of the context-object associations as determined by our third validation study. This regressor was z-scored separately within predictive contexts and non-predictive contexts but not separately within each subcondition so that differences in context-object associations between affective and neutral contexts were accounted for, but that differences between predictive and non-predictive conditions remained; finally, we convolved it with the hemodynamic response function. The six motion parameters were also included as additional nuisance regressors.

A temporal high-pass filter (cut-off of 128 s) was used to remove low-frequency drifts, and temporal autocorrelation across scans was modelled with an AR(1) process. Contrasts were then computed at the subject level and used for group analyses using one-sample *t*-tests. All voxels inside the brain were analyzed; we maintained the familywise error rate of $p < .05$, two-tailed, at the cluster level (primary threshold of $p < .001$, uncorrected) using random field theory (Friston, Worsley & Frackowiak, 1994). The Anatomy (Eickhoff et al., 2005) and WFU-PickAtlas (Maldjian, Laurienti, Kraft & Burdette, 2003) toolboxes were used to identify activated brain regions based on peak Montreal Neurological Institute (MNI) coordinates.

Results

Behavioral results

Mean accuracy was 97.1% (σ = 2.8%) for the Pred-Neut condition, 97.2% (σ = 2.1%) for the Pred-Aff condition, 96.9% (σ = 2.6%) for the noPred-Neut condition, 95.0% (σ = 2.5%) for the noPred-Aff condition and 95.7% (σ = 2.4%) for the noCont condition. When comparing Pred, noPred and noCont together (ANOVA, $n = 17$), there was no effect of condition on accuracy ($F(2,16) = 2.17$, $p = .13$, $\eta^2_p = 0.12$). When comparing all conditions except noCont together (ANOVA, $n = 17$), there was a significant main effect of predictive value ($F(1,16) = 10.36$, $p = .005$, $\eta^2_p = 0.13$), a marginally significant main effect of affective value ($F(1,16) = 4.42$, $p = .052$, $\eta^2_p = 0.08$), and a marginally significant interaction between affective and predictive values ($F(1,16) = 3.97$, $p = .063$, $\eta^2_p = 0.10$).

Mean response time was 640 ms (σ = 121 ms) for the Pred-Neut condition, 650 ms (σ = 120 ms) for the Pred-Aff condition, 638 ms (σ = 110 ms) for the noPred-Neut

condition, 636 ms (σ = 127 ms) for the noPred-Aff condition and 632 ms (σ = 116 ms) for the noCont condition. When comparing Pred, noPred and noCont together (ANOVA, n = 17), there was no effect of condition on response time ($F(2,16) = 1.53$, $p = .23$, $\eta^2_p = 0.09$). When comparing all conditions except noCont together (ANOVA, n = 17), there was no main effect of affective or predictive value and no interaction (Fs(1,16) = 1.67, 0.25 and 0.64 respectively, $p > .20$, $\eta^2_p < .03$).

fMRI results

To investigate the potential effect of the generation of explicit contextual expectations (occurring only in the Pred and noPred conditions) on brain activity, we contrasted the Pred and noPred conditions with the noCont condition (paired t-test, n = 17). Five clusters were significantly more activated in the Pred and noPred conditions than in

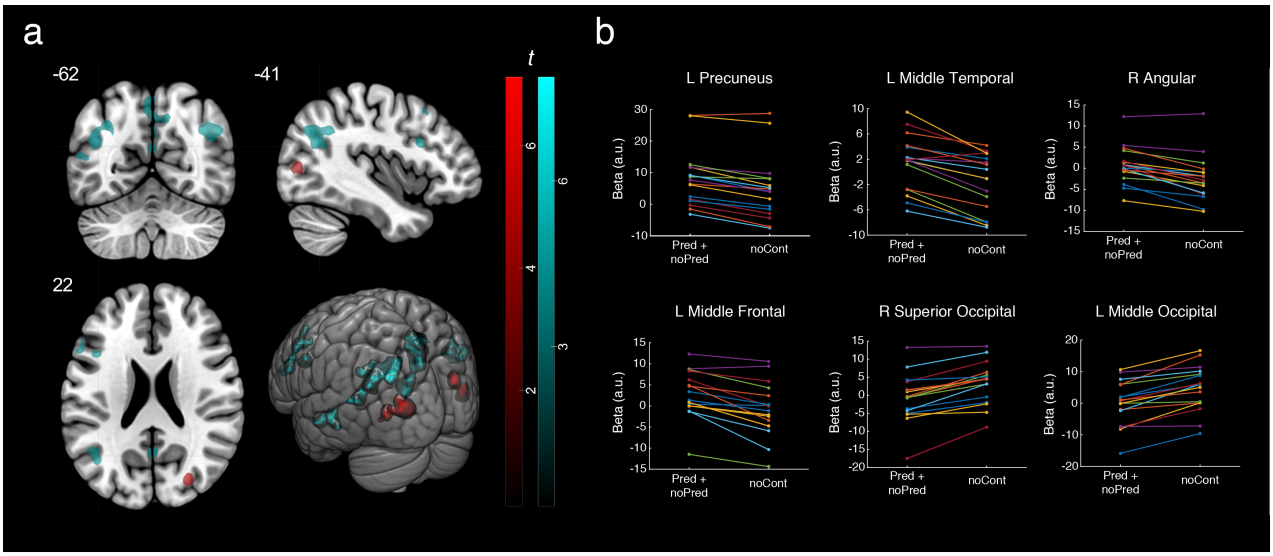


Figure 2. A) Significant clusters for the noCont > (Pred + noPred) (in red) and the (Pred + noPred) > noCont (in cyan) contrasts. **B)** Beta values of individual subjects for noCont and (Pred + noPred) conditions in peak voxels of various significant clusters.

the noCont condition (p < .05, two-tailed, corrected for family-wise error rate (FWER); peak Cohen's d_z = 1.91; Figure 2; Table 1): one bilateral cluster in the precuneus, one extending from the left precuneus and middle occipital gyrus to the left angular gyrus, one in the left middle temporal gyrus, one in the left middle and inferior frontal gyri and one in the right angular gyrus. The reverse contrast revealed the specific activation of two clusters in the right superior and middle occipital gyri and in the left middle occipital gyrus (p < .05, two-tailed, FWER-corrected; peak Cohen's d_z = 1.73; Figure 2; Table 1).

We then investigated whether there was a main effect of predictive value (Pred vs noPred), a main effect of the context's affective value (Aff vs Neut), and an interaction between predictive and affective values on brain areas involved in object recognition (paired t-tests, n = 17). There were no significant main effects of predictive and affective values. However, there was a significant interaction between predictive and affective values for two clusters: one in the right cuneus and one overlapping the left cuneus, calcarine gyrus and lingual gyrus (p < .05, two-tailed, FWER-corrected; peak Cohen's d_z = 1.74; Figure 3; Table 1). We then investigated what simple effects resulted in this interaction: when looking at the simple effects on the peak voxels of each significant cluster, we observed that they were more active in the Pred-Aff condition than in the

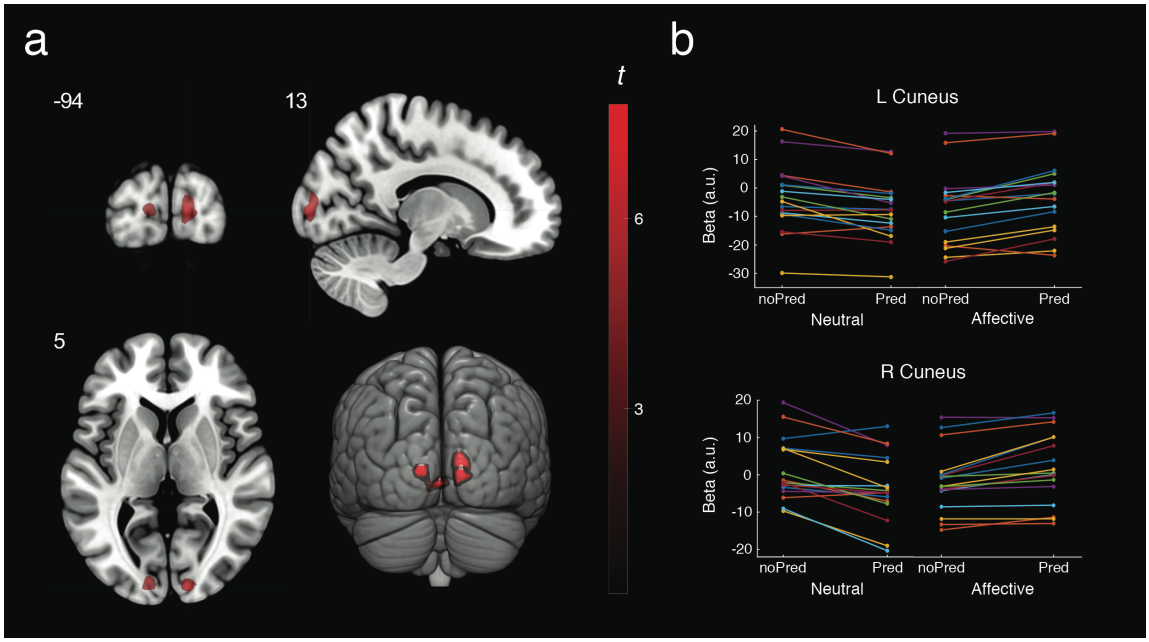


Figure 3. A) Significant clusters for the interaction between affective and predictive values (in red). **B)** Beta values of individual subjects for each condition in peak voxels of significant clusters.

noPred-Aff condition (left cuneus: $t(16) = 4.80$, $p_{Bonf} = .0008$, $d_z = 1.16$; right cuneus: $t(16) = 4.56$, $p_{Bonf} = .001$, $d_z = 1.11$) and more active in the noPred-Neut than in the Pred-Neut condition (left cuneus: $t(16) = 4.41$, $p_{Bonf} = .002$, $d_z = 1.07$; right cuneus: $t(16) = 4.24$, $p_{Bonf} = .003$, $d_z = 1.03$).

Next, we conducted a series of control analyses to ensure that the interaction could not have been the result of undesirable confounds. First, we investigated whether the interaction could have been caused by differences between the objects associated to neutral contexts and those associated to affective contexts by assessing if there was any significant difference in brain activity when they were perceived without a context (noCont condition). There was no significant difference between the conditions ($p_{FWER} > .33$). We also analyzed the image similarities directly: we used the HMAX model (Riesenhuber & Poggio, 1999; Serre, Wolf, Bileschi, Riesenhuber & Poggio, 2007), a commonly used model of the early visual cortex, and we computed correlation distances between the responses of the model to each image. We then verified if the between-categories (affective context objects to neutral context objects) distances were larger than the within-categories distances (two sample t-tests): no difference was observed (compared to within-neutral distances: .498 vs .504, $t(3430) = .57$, $n = 1128$ and 2304, $p = .57$; compared to within-affective distances: .498 vs .502, $t(3430) = .38$, $n = 1128$ and 2304, $p = .70$), suggesting that the images in these two object categories are similar.

Finally, the possibility remained that attention could explain the interaction between affective and predictive values: a similar interaction has indeed been previously reported with attention as a factor instead of affective value (Kok et al., 2012b). A first objection to this claim would be that our behavioral results actually point to an opposite effect: although we observe the same reversed prediction effect for affective contexts that Kok et al. observed for task-relevant stimuli, the lower recognition accuracy in the affective condition suggests that they are not attended more and that attention is not the cause of this interaction. Nonetheless, we decided to conduct an additional behavioral experiment to isolate potential attentional effects better. Twenty-four participants performed a Gabor orientation discrimination task (vertical vs horizontal), in which the Gabor patches (1 cycle per degree) were randomly following either a neutral context image or an affective context image in the same way as in the fMRI experiment (contexts presented for 1s, 1.5-4 s jitter,

patches presented during 133 ms); adaptive procedures were conducted separately in each condition in order to find the contrast sensitivity threshold associated with each condition. Again, no difference was observed ($\log_{10}$(contrast) of -2.10 vs -2.11; p = .94). Since we know that contrast sensitivity is greatly enhanced by attention (see Carrasco, 2006, for a review), it does not seem likely that affective contexts were attracting attention and maintaining it for up to 4s in order for it to alter object processing.

Table 1. Montreal Neurological Institute (MNI) coordinates and T values for significantly activated brain regions.

Brain regions	**Peak MNI coordinates**			**Nb of voxels**	**Peak T value**
	x	**y**	**z**		
(Pred + noPred) > noCont					
L Middle Occipital	-30	-69	39	232	7.88
L Angular	-42	-63	27		6.63
L Middle Temporal	-54	-21	-9	76	6.92
R Angular	51	-63	36	90	6.16
L Middle Frontal	-48	18	39	94	5.79
L Inferior Frontal	-42	15	24		5.51
L Precuneus	-3	-66	54	224	5.59
R Precuneus	9	-57	42		5.32
noCont > (Pred + noPred)					
R Superior Occipital	27	-78	21	61	7.12
R Middle Occipital	36	-81	12		4.75
L Middle Occipital	-27	-84	6	90	6.76
Predict. x Affect.					
R Cuneus	15	-96	6	45	7.79
L Cuneus	-12	-90	3	43	6.08
	0	-87	-3		4.55
	-9	-84	-9		3.89

Discussion

Our first aim was to investigate how the generation of expectations about objects from a preceding context might modulate the activity of brain areas involved in object perception. We found significantly more activation in the precuneus, the left middle occipital gyrus, the left middle temporal gyrus, the left frontal cortex and the parietal cortex, when (valid or invalid) contextual expectations were generated prior to object perception, suggesting that these high-level areas are mainly associated with object processing when expectations are generated. These activations specifically represent an interaction between contextual expectations and object bottom-up sensory information: activity related solely to object processing is cancelled out because the objects are the same in both conditions, and activity related solely to the prior presentation of the context is regressed out in the GLM.

To our knowledge, only Summerfield & Koechlin (2008) performed a similar analysis before; however, they used lines as cues and gratings as stimuli, and the cue was directly related to the task (the subjects had to indicate whether the cue and the grating matched). In their study, they observed a significantly greater activation of the middle occipital and fusiform gyri when there was an expectation. We also find a greater activation of the middle occipital gyrus, in addition to many other brain regions. Since expectations in our study are about objects rather than simple grating orientations, regions representing them are likely to be more numerous. The interaction between object and context processing observed in the middle temporal gyrus (a part of the inferotemporal cortex) supports a popular hypothesis according to which top-down contextual predictions would be combined with bottom-up sensory information to facilitate object recognition in the inferotemporal cortex (Bar, 2004). The precuneus and the parietal cortex, which are also activated in this contrast, have previously been linked to episodic memory retrieval and contextual associative processing (Lundstrom, Ingvar & Petersson, 2005; Aminoff, Gronau & Bar, 2007; Livne & Bar, 2016; Brandman & Peelen, 2017) which both require the integration of stored representations with incoming sensory information. Moreover, the precuneus of an observer that views several objects simultaneously is more activated when these objects are contextually related than when they are not (Livne & Bar, 2016); this

suggests that the contextual representations elicited by some of these objects are compared to other objects. Recently, activity in the retrosplenial complex, a region comprising the precuneus, has been shown to correlate with supra-additive decoding of objects embedded in scenes, suggesting that the precuneus is responsible for a scene-based facilitation of object representations (Brandman & Peelen, 2017). Interestingly, the interaction we observed between context and object information in the precuneus is also supra-additive (i.e. there is a remaining positive activation after considering the main effects of object and context). We extend previous results by showing that the precuneus integrates object sensory information with valid or invalid scene-based expectations generated prior to object presentation. The inferior and middle frontal gyri were also active during object processing when expectations were generated. These regions have previously been found to respond more to objects in non-congruent scenes than to objects in congruent scenes (Rémy et al., 2014): it is thus likely that they are responsible of integrating contextual information with perceived objects. Other frontal areas have previously been found to both maintain expectations and integrate them with sensory information (Summerfield et al., 2006; Summerfield & Koechlin, 2008).

When investigating which regions were decoding objects *in* scenes better than objects *and* scenes (in a supra-additive manner), Brandman & Peelen (2017) reported lateral extrastriate loci of activations, including the lateral occipital cortex and the posterior fusiform sulcus. These regions largely differ from the ones we uncovered (most notably the precuneus and the frontal cortex), suggesting that the matching of automatic contextual expectations with sensory evidence recruits different regions than the ones involved in the simultaneous integration of object and background. This result implies that these two processes may be distinct.

The reverse contrast, associated with visual processing of objects when no expectation (neither valid nor invalid) had been generated from a context, yielded bilateral activation of primary visual areas. Activated voxels may be part of areas primarily associated with the processing of sensory information shared by a majority of objects (e.g., intermediate spatial frequencies; Caplette, West, Gomot, Gosselin & Wicker, 2014), which is thus reduced when almost any object is expected.

We then investigated whether there was an effect of prediction error or match, i.e. whether some areas were more active at the presentation of the object when the object followed a predictive context or when the object followed a non-predictive context. When neutral and affective contexts were combined, there was no significant difference between predictive and non-predictive conditions; however, there was a significant interaction between predictive and affective values in low-level occipital areas, specifically the left and right cunei. Looking at these clusters, the classical prediction error effect was visible for neutral contexts, i.e. predicted objects elicited a smaller BOLD signal; but, when contexts were affective, this effect was reversed, i.e. predicted objects elicited a larger BOLD signal. Note that previous studies observing a smaller signal for predicted objects have exclusively used affectively neutral cues, making our results compatible with theirs. Furthermore, these brain regions are different from those responding differentially to congruent and incongruent context-object pairs, typically higher-level regions such as the lateral occipital and frontal cortices (e.g., Jenkins et al., 2010; Rémy et al., 2014). This further indicates that scene-object interactions and scene-based expectations are different processes.

These results are not compatible with the proposal that a subject's internal affective state is altering the content of their predictions about object identities (Barrett & Bar, 2009). According to this idea, the affective value of a preceding context (or even a simultaneous context or the object itself; see Barrett & Bar, 2009) would alter the subject's bodily state and bring additional information that could be used by the brain to predict the identity of perceived objects. Consequently, a similar pattern of results should be visible for neutral and emotional contexts, with only a greater difference in activation between predicted and unpredicted objects for emotional contexts than for neutral contexts (due to the additional emotional information).

Our results are compatible, however, with the general idea that affect interacts with predictive processing (Barrett & Simmons, 2015; Miller & Clark, 2018). One possibility recently put forward by some authors is that, rather than contributing to the content of the predictions, a subject's internal affective state modulates the precision of the predictions (Miller & Clark, 2018). In recent formulations of predictive coding (Feldman & Friston, 2010), the prediction error is weighted by the reliability, or precision, of sensory

information. When precision is low, prediction errors are down-weighted and observers rely more on predictions; when it is high, prediction errors are up-weighted and observers rely more on the sensory input. Because this weighting only occurs for neurons representing prediction error and not for neurons representing predictions (Friston, 2009; Kok et al., 2012b), it should lead to an interaction between prediction error and precision (Rao, 2005; Friston, 2009; Kok et al., 2012b), exactly like the one we observed between predictive and affective values.

Kok and colleagues (Kok et al., 2012b) reported a similar reversal of the prediction error effect in the early visual cortex for task relevant stimuli. They argued that this effect was caused by endogenous attention enhancing the precision of the predictions (Rao, 2005; Feldman & Friston, 2010). This cannot be the cause of the effect we observed however, since attention was not manipulated in our study and our stimuli were all similarly task relevant. Furthermore, exogenous attention also similar between our conditions, as revealed by behavioral results obtained in the scanner and in the control contrast sensitivity experiment. This implies that the prediction error reversal in our study was not caused by an increase in attention.

In summary, real-world expectations initiated by contexts, irrespectively of their degree of validity, led to more activation of high-level areas (including parietal and frontal cortices) during subsequent object recognition; notably, these regions were distinct from those responsible of instantaneous scene-object interactions. Furthermore, the context's affective value interacted with the validity of the prediction it had initiated: classical prediction error effects were only observed with neutral contexts, and a complete reversal of these effects was observed when contexts were emotional. This result is not compatible with the idea that the affective value of a stimulus, and the ensuing internal bodily state of the subject, are contributing to the creation of predictions (Barrett & Bar, 2009); but it is compatible with a modulatory role of affective value over the weight of predictions in perception (Miller & Clark, 2018). In conclusion, our results deepen our understanding of predictive coding in an ecological setting by showing that the mere presence of explicit expectations, and their affective content, modulate object recognition.

References

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience, 30*(8), 2960–2966.

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex*, *17*(7), 1493–1503.

Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*(2), 347–358.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America, 103*(2), 449–454.

Barrett, L. F., & Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1325–1334.

Barrett, L. F., Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*, 419-429.

Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience*, *37*(32), 7700–7710.

Caplette, L., West, G., Gomot, M., Gosselin, F., & Wicker, B. (2014). Affective and contextual values modulate spatial frequency use in object recognition. *Frontiers in Psychology*, *5*:512.

Carrasco, M. (2006). Covert attention increases contrast sensitivity: psychophysical, neurophysiological and neuroimaging studies. *Progress in Brain Research*, *154A*, 33–70.

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 1–16.

den Ouden, H. E. M., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience, 30*(9), 3210–3219.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage, 25*(4), 1325–1335.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*:215.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. B: Biological sciences, 360*(1456), 815–836.

Friston, K. J. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301.

Friston, K. J., Worsley, K. J., & Frackowiak, R. S. J. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping, 1*, 214-220.

Goh, J. O. S., Siong, S. C., Park, D., Gutchess, A., Hebrank, A., & Chee, M. W. L. (2004). Cortical areas involved in object, background, and Object-Background Processing Revealed with Functional Magnetic Resonance Adaptation. *Journal of Neuroscience*, *24*(45), 10223–10228.

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience, 20*(3), 371–388.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665–667.

Jenkins, L. J., Yang, Y. J., Goh, J., Hong, Y. Y., & Park, D. C. (2010) Cultural differences in the lateral occipital complex while viewing incongruent scenes. *Social Cognitive Affective Neuroscience, 5*, 236–241.

Jiang, J., Summerfield, C., & Egner, T. (2013). Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *Journal of Neuroscience, 33*(47), 18438–18447.

Kirk, U. (2008). The neural basis of object-context relationships on aesthetic judgment. *PLoS ONE, 3*(11):e3754.

Kok, P., Failing, M. F., & de Lange F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience, 26*(7), 1546–1554.

Kok, P., Jehee, J. F. M., & de Lange F. P. (2012a). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2) , 265–270.

Kok, P., Mostert, P., & de Lange F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America, 114*(39), 10473–10478.

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012b). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex, 22*(9):2197–2206.

Kok, P., & Turk-Browne, N. B. (2018). Associative prediction of visual shape in the hippocampus. *Journal of Neuroscience, 38*(31):6888–6899.

Lebrecht, S., Bar, M., Barrett, L. F., Tarr, M. J. (2012). Micro-valences: perceiving affective valence in everyday objects. *Frontiers in Psychology, 3*:107.

Livne, T, & Bar, M (2016). Cortical integration of contextual information across objects. *Journal of Cognitive Neuroscience, 28*(7):948–958.

Lundstrom, B. N., Ingvar, M., Petersson, K. M. (2005). The role of preuneus and left inferior frontal cortex during source memory episodic retrieval. *Neuroimage 27*(4), 824–834.

Maldjian, J. A., Laurienti, P. J., Kraft R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage, 19*(3), 1233–1239.

Miller, M., & Clark, A. (2018). Happily entangled: prediction, emotion, and the embodied mind. *Synthese, 195*(6), 2559–2575.

Mumford, D. (1992). On the computational architecture of the neocortex: II. The role of cortico-cortical loops. *Biological Cybernetics, 66*:241.

Rao, R. P. N. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport, 16*(16), 1843–1848.

Rao, R. P. N., &Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87.

Rémy F., Vayssière N., Pins D., Boucart M., & Fabre-Thorpe M. (2014). Incongruent object/context relationships in visual scenes: Where are they processed in the brain? *Brain and Cognition, 84*(1), 34–43.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Serre, T., Wolf, L, Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*, 411–426.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science, 314*(5803), 1311–1314.

Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron, 59*(2), 336–347.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience, 11*(9), 1004–1006.

Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience, 32*(39), 13389–13395.

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex, 5*(1), 1–11.

Acknowledgements

This study was funded by the Fondation Planiol (BW), by the Institut Universitaire de France (MM) and by the Social Sciences and Humanities Research Council of Canada (LC).

Supplementary Data

Table S1. Objects associated to each context.

Affective Contexts		**Associated objects**	
Cemetery	Tombstone	Wreath	Memorial plaque
Funfair	Cotton candy	Bumper car	Ferris wheel
Fire	Fire truck	Fire hose	Fire extinguisher
Luxury Hotel	Crystal chandelier	Large bed	Champagne bottle
Ski Slope	Skis	Chair lift	Ski poles
Dumping ground	Plastic bag	Old shoe	Garbage truck
Circus	Clown nose	Trapeze	Hoop
Beach	Beach towel	Parasol	Beach ball
Nursery	Teddy bear	Mobile	Changing mat
Stage	Electric guitar	Microphone	Speakers
Wedding	Wedding cake	Wedding dress	Wedding bouquet
War zone	Machine gun	Military helmet	Grenade
Birthday	Birthday cake	Present	Birthday candles
Nightclub	Disco light	Disco ball	Cocktail glass
Hospital room	Syringe	Stethoscope	Surgical mask
Boat	Life buoy	Rudder	Life vest
Neutral Contexts			
Hairdresser	Scissors	Comb	Hairdryer
Office	Laptop computer	Pen	Stapler
Farm	Tractor	Combine	Fork
Street	Car	Road sign	Traffic lights
Bathroom	Towel	Soap	Sink
Kitchen	Pot	Oven	Refrigerator
Supermarket	Shopping cart	Shopping basket	Cash register
Garage	Damaged car	Adjustable wrench	Tire
Classroom	Black board	Chalks	Pencil case
Church	Church bench	Crucifix	Altar
Construction site	Site helmet	Crane	Shovel
Airport	Luggage trolley	Luggage	Airport bench
Plane	Plane seats	Airport window	Airplane reactor
Café/Bar	Beer glass	Coffee cup	Bar table
Tennis court	Tennis racket	Tennis ball	Tennis net
Swimming pool	Swimsuit	Diving board	Pool ladder