Influence des caractéristiques perceptives et émotionnelles des expressions faciales dans la programmation de saccades oculaires

Léa Entzmann

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement
Spécialité : PCN - Sciences cognitives, psychologie et neurocognition
Unité de recherche : Laboratoire de Psychologie et Neuro Cognition

Influence des caractéristiques perceptives et émotionnelles des expressions faciales dans la programmation de saccades oculaires

The influence of perceptual and emotional characteristics of facial expressions on saccade programming

Présentée par :

Léa ENTZMANN

Direction de thèse :

Martial MERMILLOD PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Directeur de thèse
Nathalie GUYADER habilitation ponctuelle 2015-2018, UJF	Co-encadrante de thèse

Rapporteurs :

VALERIE GOFFAUX
Professeur, Université catholique de Louvain
THERESE COLLINS
Professeur des Universités, UNIVERSITE DE PARIS-CITE

Thèse soutenue publiquement le **22 septembre 2022**, devant le jury composé de :

MARTIAL MERMILLOD Professeur des Universités, UNIVERSITE GRENOBLE ALPES	Directeur de thèse
VALERIE GOFFAUX Professeur, Université catholique de Louvain	Rapporteure
THERESE COLLINS Professeur des Universités, UNIVERSITE DE PARIS-CITE	Rapporteure
MONICA BACIU Professeur des Universités, UNIVERSITE GRENOBLE ALPES	Présidente
DIMITRI BAYLE Maître de conférences, UNIVERSITE PARIS 10 - NANTERRE	Examinateur

Invités :

NATHALIE GUYADER
Maître de conférences, UNIVERSITE GRENOBLE ALPES

Résumé

Les expressions faciales sont des stimuli visuels complexes, caractérisés à la fois par une configuration spatiale spécifique et une émotion qu'ils communiquent. Leur détection est essentielle, que ce soit dans le cadre de la survie ou de la génération de comportements sociaux adaptés. Certains modèles du traitement des émotions suggèrent que les visages avec une expression émotionnelle, en particulier apeurée, sont détectés rapidement, et ce indépendamment de l'objectif de l'observateur. Cette détection rapide ferait intervenir une voie sous-corticale, qui relie le colliculus supérieur à l'amygdale et qui traiterait uniquement l'information visuelle grossière, transmise par les basses fréquences spatiales. Cependant, toutes les recherches ne s'accordent pas forcément. L'objectif de ce travail de thèse était de préciser les processus par lesquels les expressions faciales émotionnelles (joyeuses ou apeurées) influencent la programmation de saccades oculaires, en comparaison à des expressions faciales neutres. Nous avons en particulier testé l'hypothèse d'une détection rapide ($<$ 100 ms) et indépendante de l'objectif des participants, qui favoriserait l'orientation du regard vers des visages émotionnels, particulièrement apeurés. Nous supposions que cet effet soit originaire du traitement des basses fréquences spatiales au sein de la voie sous-corticale. À travers une série d'études en choix saccadique, les résultats d'un premier chapitre expérimental témoignent d'un traitement privilégié des visages émotionnels, en particulier joyeux, qui capturent le regard plus efficacement que les visages neutres. Cependant, cet effet ne serait pas automatique, mais plutôt dépendant de la tâche de l'observateur. Nous avons aussi observé de manière systématique des différences entre les points d'arrivée des saccades sur des visages joyeux et des visages apeurés. Plus précisément, les saccades arrivaient plus près de la bouche pour les visages joyeux qu'apeurés. Les résultats d'un second chapitre expérimental témoignent de l'importance de l'information transmise par les hautes fréquences spatiales dans la détection et le déclenchement d'une saccade vers des visages neutres et émotionnels. À l'aide d'un réseau de neurones convolutionnel, nous avons pu mettre en évidence la région la plus diagnostique à cette tâche : la bouche. Les résultats d'un troisième chapitre expérimental, basé sur une étude en neuroimagerie, témoignent d'une sensibilité aux expressions faciales dans des régions corticales, indépendamment des fréquences spatiales. Cependant, nous n'avons pas observé d'effet des expressions faciales dans les régions qui constituent la voie sous-corticale. Dans la plupart des régions étudiées, les réponses neuronales étaient plus fortes face à des visages présentés en hautes plutôt qu'en basses fréquences spatiales. Les résultats de ce troisième chapitre expérimental sont discutés en lien avec les statistiques de nos stimuli. Ainsi, les travaux menés dans le cadre de cette thèse, qui allient comportement, neuroimagerie et modélisation, suggèrent que les visages émotionnels peuvent attirer le regard plus efficacement que les visages neutres. Cependant, cet effet ne serait pas automatique. Nous proposons que, en particulier dans les tâches qui exigent une réponse rapide, les visages émotionnels ne vont pas attirer l'attention et le regard plus que les visages neutres. Mais, les caractéristiques physiques des expressions vont moduler l'attention automatiquement, ce qui se traduit par des décalages au niveau des points d'arrivée des saccades. Finalement, que ce soit au niveau comportemental ou neural, nos résultats n'ont pas permis de mettre en avant un traitement privilégié des visages apeurés, basé sur l'extraction rapide des basses fréquences spatiales. Ils remettent ainsi en question l'implication de la voie sous-corticale dans le traitement des expressions faciales.

Mots-clés : Expressions faciales émotionnelles ; Mouvements oculaires ; Attention visuelle ; Fréquences spatiales ; Réseaux de neurones artificiels ; IRMf

Abstract

Facial expressions are complex visual stimuli, characterized by both a specific spatial configuration and the emotion they communicate. Their detection is essential, whether in the context of survival or in the generation of adapted social behaviours. Some models of emotional processing suggest that faces with emotional expressions, especially fearful ones, are detected rapidly, regardless of the observer's goal. This rapid detection is thought to involve a subcortical pathway, which connects the superior colliculus to the amygdala and processes only coarse visual information, transmitted through low spatial frequencies. However, not all research is in agreement. The aim of this thesis work was to clarify the processes by which emotional facial expressions (happy or fearful) influence the programming of saccadic eye movements, compared to neutral facial expressions. In particular, we tested the hypothesis of a rapid (< 100 ms) and task-independent detection, which would favour gaze orientation towards emotional, particularly fearful, faces. We hypothesised that this effect originated from low spatial frequency processing within the subcortical pathway. Through a series of saccadic choice experiments, the results of a first experimental chapter show a privileged processing of emotional faces, in particular happy ones, which capture the gaze more efficiently than neutral faces. However, this effect was not automatic, but rather dependent on the observer's task. We also systematically observed differences between saccade endpoints on happy and fearful faces. Specifically, saccades landed closer to the mouth for happy faces than for fearful faces. The results of a second experimental chapter demonstrate the importance of high spatial frequency information in the detection and initiation of saccades toward neutral and emotional faces. Using a convolutional neural network, we were able to identify the most diagnostic region for this task : the mouth. The results of a third experimental chapter, based on a neuroimaging study, show sensitivity to facial expressions in cortical regions, independent of spatial frequencies. However, we did not observe an effect of facial expressions in regions that constitute the subcortical pathway. In most of the regions studied, neural responses were stronger to faces presented at high rather than low spatial frequencies. The results of this third experimental chapter are discussed in relation to the statistics of our stimuli. Thus, the work conducted in this thesis, which combines behaviour, neuroimaging and modelling, suggests that emotional faces can attract the gaze more effectively than neutral faces. However, this effect would not be automatic. We propose that, particularly in tasks that require a quick response, emotional faces will not attract attention and gaze more than neutral faces. However, the physical characteristics of the expressions will modulate attention automatically, resulting in shifts in saccade endpoints. Finally, whether at the behavioural or neural level, our results did not reveal a privileged processing of fearful faces, based on the rapid extraction of low spatial frequencies. They thus questionned the involvement of the subcortical pathway in the processing of facial expressions.

Keywords : Emotional facial expressions ; Eye movements ; Visual attention ; Spatial frequencies ; Artificial neural netwoks ; fMRI

Remerciements

Ce manuscrit est le fruit de quatre années de travail qui se sont révélées intenses, mais surtout enrichissantes aussi bien d'un point de vue professionnel que personnel. Je souhaite ici remercier toutes les personnes qui, de près ou de loin, ont pu contribuer à son aboutissement.

Je tiens à adresser mes premiers remerciements à mes encadrants de thèse, Martial Mermillod et Nathalie Guyader, sans qui ce projet n'aurai sans doute jamais vu le jour. Travailler avec vous fut un réel plaisir. Merci à Nathalie de m'avoir initiée à la recherche lors de ma première année de master. Je pense que je partais d'assez loin, et pourtant vous m'avez toujours accompagnée, jusqu'en thèse. Merci pour la rigueur scientifique que vous avez su me transmettre et votre soutien tout au long de ces années. Martial, merci d'avoir cru en moi tout au long de ce projet de thèse. Je vous remercie tous les deux pour votre confiance, votre optimisme, vos encouragements, vos précieux conseils et vos relectures.

Ensuite, je remercie très chaleureusement l'ensemble des membres du jury, Monica Baciu, Thérèse Collins, Valérie Goffaux et Dimitri Bayle, qui m'ont fait l'honneur de lire et d'évaluer mon travail. Je vous remercie pour votre bienveillance et vos retours précieux. Je remercie également Anne Guérin-Dugué et Dimitri Bayle pour leur accompagnement lors des comités de suivi individuel de thèse.

Merci à l'Université Grenoble Alpes (bourse IDEX - IRS) d'avoir financé mes recherches.

Je remercie sincèrement Louise Kauffmann et Carole Peyrin pour leur accompagnement tout au long de ces travaux. Votre expertise et vos retours sur les expériences ainsi que sur les articles ont largement contribué à la richesse de ces travaux. Merci pour votre temps et votre bienveillance.

Merci également à Juliette Lenouvel, Clémence Charles et Céline Michel pour leur implication dans ce projet dans le cadre de leur projet de master. Ce fut très agréable pour moi de travailler avec vous.

Merci à Roman Vuillaume pour son aide sur la partie réseaux de neurones artificiels.

Un grand merci à Émilie Cousin pour son temps et son aide sur le traitement des données IRM. Merci pour ta bienveillance, tes explications claires et ta rigueur. Ces données nous ont parfois donné du fil à retordre, mais c'était un plaisir de travailler avec toi.

Merci aux membres du projet EyeProxy, Michel Dojat, Anne Guérin-Dugué, Gaëlle Nicolas et Emmanuelle Kristensen de m'avoir permis d'intégrer ce projet ambitieux et passionnant. C'était un plaisir de collaborer avec vous. Ce projet a été très enrichissant

pour moi et m'a permis d'en apprendre beaucoup sur la neuroimagerie. Je remercie particulièrement Gaëlle Nicolas pour sa compagnie lors des séances d'acquisition. J'ai beaucoup apprécié partager un bout de ton projet de thèse.

Je remercie aussi chaleureusement l'ensemble du LPNC, qui m'a fourni un environnement de travail des plus agréable. C'est une chance d'effectuer une thèse dans un environnement aussi stimulant et bienveillant.

Je remercie (évidemment) particulièrement mes collègues et amis du bureau E121, Adeline, Cynthia, Elie, Sarah, Wilfried, Candice, Gull et David. L'ambiance au bureau était incroyable, je ne l'oublierai jamais. C'était super d'avoir pu partager mes journées avec vous, entre fous rires, doutes et décoration d'intérieur. Je suis très heureuse d'être tombée dans le même bureau que vous et j'espère que nous aurons l'occasion de nous revoir. Merci Audrey, membre non-officielle du bureau E121, pour ta gentillesse et ton soutien tout au long de ces années. Ce fut également un plaisir d'évoluer à tes côtés. Merci Merrick pour les séances de badminton (suivies de manière assidue) et ta bienveillance. De manière générale merci aux doctorants du LPNC pour leur dynamisme et leur enthousiasme : Olivier, Méline, Rémi, Lucie, Lise, Pauline, Laura, Lucrèce, Maëlle, Brice, Élise, Célise, Ali, Samuel et tous les autres...

Merci à mes amies de master, Flora et Sonja, pour leurs encouragements, leur écoute et les discussions que nous avons partagées. Merci pour les soirées soupes et jeux de société qui vont me manquer. Vous avez été une grande source d'inspiration pour moi.

Merci à ma famille pour leur soutien au quotidien.

Pour terminer, je remercie Gabriel, qui me supporte et me soutient depuis maintenant plusieurs années. Merci d'avoir rendu mes journées plus agréables, j'ai hâte de voir où le futur nous emmène.

Table des matières

Liste des figures

Liste des tableaux

Liste des abréviations

AAL . . . . . *Automated Anatomical Labeling.*
BFS . . . . . Basses fréquences spatiales.
CGL . . . . . Corps genouillé latéral.
CNN . . . . . Réseau de neurones convolutionnel ou *convolutional neural network.*
cpd . . . . . Cycles par degré d'angle visuel.
cpi . . . . . Cycles par image.
CS . . . . . Colliculus supérieur.
DCM . . . . . Modélisation causale dynamique ou *dynamic causal modeling.*
EG . . . . . Égalisé.
EEG . . . . . Électroencéphalographie.
EFE . . . . . Expressions faciales émotionnelles.
FACS . . . . . *Facial action coding system.*
FFA . . . . . Aire fusiforme des visages ou *fusiform face area.*
HFS . . . . . Hautes fréquences spatiales.
IRMf . . . . . Imagerie par résonance magnétique fonctionnelle.
KDEF . . . . . *Karolinska Directed Emotional Faces Database.*
MEG . . . . . Magnétoencéphalographie.
MNI . . . . . *Montreal Neurological Institute.*
MLP . . . . . Perceptron multicouche ou *multilayer perceptron.*
NonEG . . . . . Non égalisé.
OFA . . . . . Aire occipitale des visages ou *occipital face area.*
OFC . . . . . Cortex orbitofrontal ou *orbitofrontal Cortex.*
PCS . . . . . Pourcentage de changement de signal.
pSTS . . . . . Sillon temporal supérieur postérieur ou *posterior superior temporal sulcus.*
ROI . . . . . Région d'intérêt ou *region of interest.*
RMS . . . . . *Root mean square.*
SPM . . . . . *Statistical parametric mapping.*
V1 . . . . . Cortex visuel primaire.

Abréviations spécifiques aux articles

BSF . . . . . *Broad spatial frequencies* (A2, A3).
EFE . . . . . *Emotional facial expressions* (A1, A2, A3).
EQ . . . . . *Equalized* (A2).
HSF . . . . . *High spatial frequencies* (A2, A3).
LSF . . . . . *Low spatial frequencies* (A2, A3).
MSE . . . . . *Mean squared error* (A1).
NonEQ . . . . . *Non-equalized* (A2).
RMS . . . . . *Relative mouth saliency* (A3).

Avant-propos

Dans la vie de tous les jours, l'humain ne cesse d'explorer visuellement son environnement. Cette exploration est rythmée par une alternance de saccades et de fixations du regard sur différents points d'intérêts. Naturellement, certains stimuli visuels vont avoir tendance à attirer notre attention et notre regard plus efficacement que d'autres. Par exemple, les visages, qui transmettent une multitude d'informations pertinentes d'un point de vue social ou évolutif. La littérature scientifique a permis de mettre en avant une détection rapide et efficace des visages, qui sont traités par un réseau neuronal vaste et distribué. Parmi toutes les informations transmises par les visages, nous nous intéressons dans ce travail de thèse à une caractéristique particulière : l'expression faciale. Plus spécifiquement, nous avons tenté de mieux comprendre comment les expressions faciales émotionnelles influencent les mécanismes impliqués dans la programmation des mouvements oculaires. Les expressions faciales émotionnelles sont des stimuli visuels complexes, caractérisés à la fois par une configuration spatiale spécifique et une émotion qu'ils communiquent. Plusieurs recherches suggèrent que les expressions faciales émotionnelles attirent particulièrement le regard en comparaison aux expressions faciales neutres. Cependant, la nature de cet effet, son décours temporel ainsi que les réseaux neuronaux impliqués ne sont pas bien connus. Certains modèles proposent que le traitement privilégié des visages émotionnels, en particulier ceux qui transmettent une expression apeurée, puisse se faire automatiquement (c'est-à-dire rapidement et indépendamment de la tâche) sur la base d'un traitement de l'information grossière. C'est l'une des hypothèses que nous allons tester dans le présent manuscrit, qui se découpe en cinq chapitres. Dans le premier chapitre, nous proposons d'aborder le contexte théorique qui a guidé nos questions de recherche. Les chapitres deux, trois et quatre sont des chapitres expérimentaux. Parfois constitués d'articles (publiés ou soumis), ils présentent les travaux que nous avons menés pour répondre à nos questions de recherche. Le dernier chapitre nous permet de faire un bilan des différents résultats obtenus, en soulignant particulièrement ce qu'ils apportent d'un point de vue théorique ou méthodologique.

Chapitre 1

Cadre théorique

Table des matières

1.1 Préface

Nous évoluons dans un monde complexe, dans lequel une multitude d'informations nous sont accessibles. Ces informations peuvent se présenter sous la forme de signaux visuels, auditifs, tactiles, gustatifs ou encore olfactifs. L'interprétation de ces différents signaux est cruciale d'un point de vue évolutif, car elle permet la génération d'un comportement adapté à la situation. La vision est le sens dominant chez l'homme, qui possède un système spécialisé dans la réception et l'interprétation des signaux visuels. Dans cette thèse, nous nous intéressons à la perception de stimuli visuels particuliers : des visages émotionnels [1]. Plus particulièrement, nous étudierons les mécanismes impliqués dans la capture de l'attention par des visages émotionnels, qu'il est difficile d'aborder indépendamment des mécanismes plus généralement impliqués dans la perception visuelle. L'objectif de ce premier chapitre est de présenter le contexte théorique dans lequel s'inscrit ce travail de thèse. Nous commencerons par revenir sur quelques bases physiologiques et neuropsychologiques de la perception visuelle, en lien avec la notion de fréquences spatiales, ainsi que sur quelques modèles de l'attention visuelle. Nous discuterons également des mouvements oculaires, car nous étudierons l'attention visuelle en analysant la position du regard. Après avoir présenté brièvement les notions nécessaires à la compréhension des travaux menés durant cette thèse, nous présenterons un état de l'art des travaux sur la perception des visages, et plus particulièrement des visages émotionnels. Ainsi, nous discuterons d'une sélection d'études comportementales qui montrent que les visages sont traités très efficacement par le système visuel. Ensuite, nous nous intéresserons aux expressions faciales, par l'intermédiaire de leur capacité à être reconnues et à attirer l'attention et le regard. Pour finir, nous reviendrons sur les bases cérébrales du traitement des expressions faciales, et discuterons des évidences et des débats qui entourent l'hypothèse d'un traitement sous-cortical des expressions faciales, qui pourrait influencer rapidement les mécanismes de programmation des mouvements oculaires.

1.2 La perception visuelle : une analyse fréquentielle

Cette section a pour but de revenir sur quelques bases du traitement de l'information visuelle, pertinentes dans le cadre de ce travail de thèse. Dans un premier temps nous aborderons l'anatomie de la rétine, dont dépend notre système visuel. Puis, nous décrirons la voie principale du traitement de l'information visuelle : la voie rétino-géniculo-striée. Nous terminerons par discuter du rôle des fréquences spatiales au sein du système visuel, en nous basant sur des évidences à la fois anatomiques et comportementales.

1. La notion de *visage émotionnel* fait ici référence à un visage dont l'expression faciale est non neutre. Avec cette distinction, nous n'excluons pas que tous les visages puissent être, par essence, des stimuli émotionnels du fait qu'ils sont pertinents d'un point de vue social et évolutif. Néanmoins, nous établissons une hiérarchie entre les visages dont l'expression est neutre, et ceux dont l'expression n'est pas neutre, qui susciteraient une réponse émotionnelle plus forte (pour plus de détails concernant la définition d'un stimulus émotionnel, voir Brosch et al., 2010 ; Grühn et Sharifian, 2016).

1.2.1 La rétine, l'entrée du système visuel

Le traitement de l'information visuelle va débuter au niveau de la rétine, qui se situe au fond de l'oeil (pour une revue détaillée de l'organisation de la rétine, voir Masland, 2001, 2012). La rétine est l'organe sensible de la vision, qui va transformer le signal lumineux en un signal nerveux. Elle se compose de plusieurs couches de cellules, représentées sur la Figure 1.1. Le signal lumineux va entrer dans la voie de traitement de l'information visuelle par les cellules situées sur la couche la plus interne de la rétine : les cellules photoréceptrices, également appelées photorécepteurs. Ce sont ces cellules qui vont pouvoir le transformer en un signal nerveux. Il existe deux types de photorécepteurs, chacun ayant des propriétés différentes. Les plus nombreux sont les bâtonnets. Environ 130 millions, ils représentent presque 95% des photorécepteurs (Joukal, 2017). Ils sont sensibles à une lumière de faible intensité, mais pas aux couleurs et deviennent rapidement saturés lorsqu'ils sont exposés à une lumière trop intense. Moins nombreux, les cônes sont les photorécepteurs qui codent l'information chromatique. Ils ont besoin d'une forte intensité lumineuse pour s'activer. Ainsi, ce sont les principaux acteurs de la vision diurne, les bâtonnets étant principalement utiles pour la vision nocturne. Il est important de souligner que les cônes et les bâtonnets ne sont pas répartis de manière uniforme dans la rétine. Le centre de la rétine, appelé la fovéa, est composé uniquement de cônes, dont le nombre diminue rapidement avec l'excentricité. En revanche, la périphérie de la rétine est essentiellement composée de bâtonnets.

Le message nerveux issu des photorécepteurs est transmis aux cellules ganglionnaires, moins nombreuses que les photorécepteurs. Nous pouvons distinguer ici trois types de cellules ganglionnaires : les cellules magnocellulaires, les cellules parvocellulaires et les cellules koniocellulaires[2] [3] (Yoonessi et Yoonessi, 2011). Les cellules magnocellulaires possèdent de larges champs récepteurs, c'est-à-dire qu'elles reçoivent l'information d'un grand nombre de photorécepteurs, ce qui implique une faible résolution spatiale et une transmission grossière de l'information. Sensibles au contraste et aux mouvements, elles bénéficient grâce à la myélinisation de leurs axones d'une haute résolution temporelle. Les cellules parvocellulaires possèdent de petits champs récepteurs, ce qui implique une haute résolution spatiale et un traitement détaillé de l'information. Elles sont sensibles aux couleurs, et transmettent l'information plus lentement que les cellules magnocellulaires, leurs axones n'étant pas myélinisés. Très nombreuses, elles sont principalement localisées au centre de la rétine (U. S. Kim et al., 2021). En résumé, les cellules magnocellulaires permettent de répondre rapidement aux stimuli en mouvement, tandis que les cellules parvocellulaires permettent la transmission détaillée des informations (Kaplan, 2004).

La disparité dans la répartition des photorécepteurs et des cellules ganglionnaires entraîne une diminution de l'acuité en vision périphérique comparativement à la vision centrale. La Figure 1.2 simule cette disparité en présentant une information grossière en périphérie de l'image et une information détaillée au centre de l'image. Cette disparité

2. Les cellules koniocellulaires étant moins connues et ayant peu d'intérêt dans le cadre de ce travail, leur rôle ne sera pas détaillé.

3. Notons néanmoins que la classification des cellules ganglionnaires est en réalité plus complexe. Pour une revue récente de la littérature, voir par exemple U. S. Kim et al. (2021).

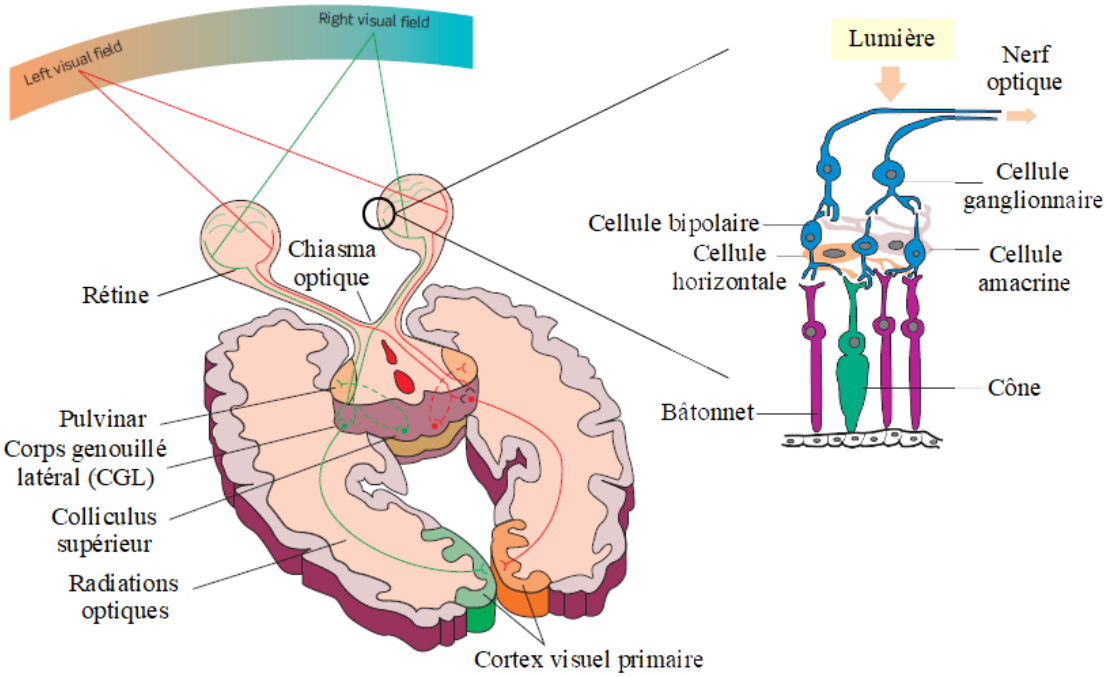

Figure 1.1 – Représentation schématique du système visuel (à gauche) et de la rétine (à droite). La voie principale du traitement visuel, la voie rétino-géniculo-striée, est représentée par les traits pleins oranges et verts. Par cette voie, l'information visuelle va des yeux au corps genouillé latéral, avant d'être transmise au cortex visuel primaire. Figure extraite de Snowden et al. (2012).

justifie notamment la programmation des mouvements oculaires, qui permet de placer une zone du champ visuel en vision centrale, où l'acuité est maximale. La sortie de la rétine se fait par l'intermédiaire des axones des cellules ganglionnaires, qui se regroupent pour former le nerf optique, la voie de sortie de l'œil.

1.2.2 La voie rétino-géniculo-striée, la voie visuelle principale

Nous appellerons ici voie rétino-géniculo-striée, ou plus simplement dans la suite de ce manuscrit voie corticale, la voie principale du traitement visuel, qui réceptionne l'information d'environ 90% des cellules ganglionnaires (voir les traits pleins oranges et verts sur la Figure 1.1). Comme son nom l'indique, elle comprend trois étapes principales : la rétine, le corps genouillé latéral (CGL) et le cortex visuel primaire, aussi appelé aire striée ou V1.

1.2.2.1 Le corps genouillé latéral

À la sortie de l'œil, la majorité des fibres du nerf optique est dirigée vers le CGL, dans la partie dorsale du thalamus. Les informations issues du champ visuel gauche, provenant de la rétine nasale de l'œil gauche et de la rétine temporale de l'œil droit (en orange sur la Figure 1.1), sont acheminées vers le CGL droit. À l'inverse, les informations

Figure 1.2 – Simulation de la perte d'acuité en vision périphérique par rapport à la vision centrale, en supposant la fixation au centre de l'image.

issues du champ visuel droit, provenant de la rétine nasale de l'œil droit et de la rétine temporale de l'œil gauche (en vert sur la Figure 1.1), sont acheminées vers le CGL gauche. Chaque CGL est formé de 6 couches de cellules : 2 couches magnocellulaires, qui reçoivent l'information des cellules magnocellulaires et 4 couches parvocellulaires, qui reçoivent l'information des cellules parvocellulaires. Il est important de préciser que, sur chacune des couches, les cellules du CGL sont réparties selon une organisation que l'on appelle rétinotopique de l'information visuelle. Cela signifie que les éléments proches les uns des autres dans la scène visuelle sont représentés sur des cellules voisines au niveau du CGL. Ainsi, les relations spatiales que les cellules ganglionnaires entretiennent entre elles au sein de la rétine sont préservées au sein du CGL, qui possède une représentation topographique de l'espace visuel (pour une description plus détaillée de l'organisation du corps genouillé latéral, voir par exemple Ghodrati et al., 2017). À la sortie du CGL, les voies parvocellulaire et magnocellulaire se poursuivent, à partir des projections des cellules du même nom, et convergent vers le cortex visuel primaire.

1.2.2.2 Les aire striée et aires extrastriées

Après le CGL, l'information est transmise au cortex visuel primaire (V1), situé aux pôles postérieurs des lobes occipitaux et constitué de six couches de cellules. C'est là que débute l'analyse de l'information visuelle. Similairement au traitement de l'information dans le CGL, un hémichamp visuel est projeté dans l'hémisphère controlatéral de V1, et les cellules de V1 s'organisent de manière rétinotopique. Ainsi, chaque cellule ne répond qu'à une petite partie de la scène, le cortex visuel primaire gauche est une carte du champ visuel droit et le cortex visuel primaire droit est une carte du champ visuel gauche (Daniel et Whitteridge, 1961 ; Wandell et al., 2007). Cependant, cette carte est déformée par rapport à la réalité. En effet, les régions du champ visuel qui sont situées en vision centrale sont surreprésentées par rapport à celles situées en vision périphérique. Ainsi, la fovéa qui code seulement 1% du champ visuel est représentée sur 50% du cortex visuel primaire. C'est ce qu'on appelle le phénomène de magnification corticale (Daniel et Whitteridge,

1961 ; Duncan et Boynton, 2003). Au sein de V1, deux types de cellules sont sensibles aux orientations et aux fréquences spatiales : les cellules simples et les cellules complexes (De Valois et al., 1982 ; Hubel et Wiesel, 1962, 2004). Elles permettraient de dresser une première ébauche des contours des objets de la scène, sous différentes résolutions. Au-delà de V1, l'information est communiquée aux aires visuelles extrastriées. Ces aires se situent au niveau du lobe occipital, pariétal ou temporal, et sont impliquées dans la suite du traitement de l'information visuelle. Elles comprennent notamment les aires V2, V3, V4 et V5, chacune spécialisée dans l'analyse de certains aspects de l'information (voir par exemple Grill-Spector et Malach, 2004). Les aires visuelles extrastriées comprennent également certaines aires spécialisées dans le traitement de catégories spécifiques comme les visages, les scènes ou les objets (voir par exemple Spiridon et al., 2006).

1.2.2.3 La voie dorsale et la voie ventrale

Il a longtemps été établi que l'information visuelle se transmet depuis V1 par l'intermédiaire de deux voies de traitement distinctes : la voie ventrale (également appelée voie du quoi) et la voie dorsale (également appelée voie du où ; Breitmeyer, 2014 ; de Haan et Cowey, 2011 ; Goodale et Milner, 1992 ; Ungerleider et Haxby, 1994). La voie ventrale serait impliquée dans la reconnaissance visuelle, et traverserait les aires V2 et V4 pour aller vers le cortex temporal. Elle serait principalement alimentée par les projections de la voie parvocellulaire. La voie dorsale serait impliquée dans la localisation spatiale des objets par rapport à l'observateur, et traverserait les aires V2, V3, V4 et V5 pour aller vers le cortex pariétal. Elle serait principalement alimentée par les projections de la voie magnocellulaire. Les premières évidences de cette dissociation viennent d'expériences comportementales effectuées avec des singes présentant des lésions du cortex inférotemporal, ou du cortex pariétal postérieur. Alors que la première lésion perturbait la reconnaissance des formes, la seconde nuisait aux performances dans des tâches nécessitant l'utilisation d'un repère spatial (Mishkin et Ungerleider, 1982). En résumé, ces deux voies traduiraient une dissociation entre un traitement pour l'action et un traitement pour la perception. Cette séparation du traitement visuel cortical en deux voies a néanmoins été remise en question par plusieurs auteurs. Dans une revue de la littérature, McIntosh et Schenk (2009) soulignent que les traitements de l'information visuelle pour l'action et pour la perception ne sont probablement pas aussi indépendants, et pourraient finalement impliquer de nombreuses interactions. D'une manière similaire, dans une autre revue de la littérature, de Haan et al. (2018) suggèrent qu'il existerait soit un plus grand nombre de voies de traitement, soit des interactions flexibles et dynamiques entre les régions de la voie dorsale et de la voie ventrale.

1.2.3 Un traitement des basses vers les hautes fréquences spatiales ?

Les paragraphes précédents ont dressé une première ébauche du système visuel, ainsi que de ses propriétés. Notamment, ils ont mis en avant un traitement parallèle et distinct de l'information grossière, en basses fréquences spatiales (BFS), et de l'information

détaillée, en hautes fréquences spatiales (HFS). En effet, les BFS sont transmises par la voie magnocellulaire et les HFS par la voie parvocellulaire ; la voie magnocellulaire permettant une transmission de l'information plus rapide que la voie parvocellulaire. Pour donner un ordre de grandeur, des enregistrements neurophysiologiques chez le singe suggèrent que l'information issue de la voie magnocellulaire atteindrait V1 environ 20 ms avant l'information issue de la voie parvocellulaire (Nowak et Bullier, 1997). De plus, les cellules de V1 sont sensibles aux fréquences spatiales. Ces propriétés, ainsi que les résultats d'études comportementales, ont conduit à l'hypothèse d'un traitement de l'information visuelle *coarse-to-fine*, basé sur la précédence du traitement des BFS sur celui des HFS (Bar, 2003 ; Hegde, 2008 ; Kauffmann, Chauvin et al., 2015 ; Kauffmann, Ramanoël et al., 2015 ; Kauffmann et al., 2014 ; Musel et al., 2014 ; Petras et al., 2019 ; Schyns et Oliva, 1994). Dans cette section, nous allons revenir sur la notion de fréquences spatiales, en présentant notamment la technique de filtrage spatial, qui permet d'isoler une bande de fréquence, afin, par exemple de tester son effet sur le comportement ou l'activité cérébrale. Ensuite, certaines données comportementales qui ont permis l'élaboration de l'hypothèse de l'analyse *coarse-to-fine*, seront présentées.

1.2.3.1 La notion de fréquences spatiales

Dans le domaine du traitement d'images, les fréquences spatiales définissent les variations de luminance par unité de distance. Elles sont mesurées en nombre d'ondulations, appelées cycles, par degré d'angle visuel (cpd) ou par image (cpi). Le nombre de cycles par image est une mesure indépendante de l'observateur, alors que le nombre de cycles par degré d'angle visuel est une mesure qui prend en compte la distance entre l'observateur et l'image. Plus il y a de cycles, que ce soit par degré ou par image, plus la fréquence spatiale est haute. Il est possible d'isoler certaines bandes de fréquences spatiales en appliquant un filtre sur une image. L'opération se fait soit dans le domaine spatial, avec l'utilisation d'un masque de convolution, soit dans le domaine fréquentiel. Sans rentrer dans les détails, la transformée de Fourier est une opération qui permet de décrire un signal dans le domaine fréquentiel comme une somme de sinusoïdes, qui se caractérise par un spectre d'amplitude et un spectre de phase. Le spectre d'amplitude décrit la distribution du contraste de luminance en fonction de la fréquence spatiale et de l'orientation. Le spectre de phase décrit la position relative des fréquences spatiales dans l'espace. Sur le spectre d'amplitude, les HFS sont représentées en périphérie, tandis que les BFS sont représentées au centre. Par exemple, en retirant la partie centrale du spectre d'amplitude, qui correspond à l'information en BFS, et en effectuant une transformée de Fourier inverse, on obtient une image en HFS. Lorsqu'une image est filtrée pour ne laisser passer que les BFS, cela se visualise sur le spectre d'amplitude par une information uniquement au centre. Cet outil mathématique est particulièrement utile à l'étude du rôle des fréquences spatiales dans la perception visuelle. En effet, cela permet d'étudier les réponses comportementales ou neurophysiologiques induites par la présentation de différentes bandes de fréquences.

1.2.3.2 Évidences en faveur d'un modèle *coarse-to-fine*

La notion de traitement *coarse-to-fine* caractérise un traitement qui débute par l'analyse de l'information grossière, en BFS, pour ensuite aller vers une analyse de l'information plus fine, en HFS. Au niveau comportemental, avant même que la notion de fréquence spatiale soit introduite, plusieurs études ont mis en avant un traitement plus rapide de l'information globale par rapport à l'information locale. Ce phénomène, connu sous le nom de précédence globale, a été introduit par Navon, en 1977, dans une étude où il utilisait des stimuli hiérarchiques. Plus précisément, des lettres globales formées de lettres locales, par exemple un grand S formé de petits H, étaient présentées à des participants. Navon a montré que la lettre globale était identifiée plus rapidement que les lettres globales. Les fréquences spatiales ont plus tard été associées à ce phénomène. Par exemple, Badcock et al. (1990) ont reproduit l'expérience de Navon en appliquant un filtrage spatial aux stimuli. Ils ont observé que, lorsque les stimuli étaient présentés en HFS, le phénomène de précédence globale disparaissait.

L'idée d'un traitement *coarse-to-fine* de l'information visuelle a été proposée dans une étude de Schyns et Oliva, en 1994. Dans cette étude, les auteurs ont utilisé des images dites hybrides, qui correspondent à la superposition d'une image en BFS et d'une image en HFS ; les deux images appartenant à des catégories différentes (Figure 1.3). Ainsi, une image hybride pouvait, par exemple contenir les BFS d'une scène d'autoroute et les HFS d'une scène de ville. Dans une première expérience, une scène hybride et une scène non filtrée étaient présentées, dans cet ordre, à des participants qui devaient répondre selon que la scène non filtrée était présente dans la scène hybride ou non. Lorsque la scène hybride était présentée brièvement (30 ms), l'interprétation était plus souvent basée sur les BFS, alors que l'inverse était observé lorsque l'image hybride était présentée plus longtemps (150 ms). Dans une deuxième expérience, deux scènes hybrides étaient présentées brièvement, l'une à la suite de l'autre, à des participants qui devaient les catégoriser. Pour chaque essai, les BFS de la première image hybride correspondaient aux HFS de la deuxième, ou inversement. Les résultats ont montré que les participants se basaient le plus souvent sur les BFS de la première image et sur les HFS de la deuxième, plutôt que l'inverse (c'est-à-dire plutôt que sur les HFS de la première image et les BFS de la deuxième). Ainsi, cette étude a mis en avant l'utilisation préférentielle d'un traitement *coarse-to-fine*, par rapport à un traitement *fine-to-coarse*. Les BFS seraient utilisées lors du traitement rapide d'une scène, tandis que les HFS seraient privilégiées lorsque la scène est traitée plus longtemps.

Plus récemment, des études réalisées au sein du Laboratoire de Psychologie et NeuroCognition ont mis en évidence cet effet en manipulation l'ordre de présentation d'images d'une même scène : des HFS vers les BFS ou l'inverse (Kauffmann, Chauvin et al., 2015 ; Kauffmann, Ramanoël et al., 2015 ; Musel et al., 2014). Ces études ont montré que la catégorisation des scènes était plus rapide lorsque la présentation se faisait des BFS vers les HFS, et donc en imposant un traitement *coarse-to-fine*. Néanmoins, plusieurs études soulignent le caractère flexible du traitement des fréquences spatiales (Morrison et Schyns, 2001 ; Oliva et Schyns, 1997 ; Ozgen et al., 2006 ; Schyns et Oliva,

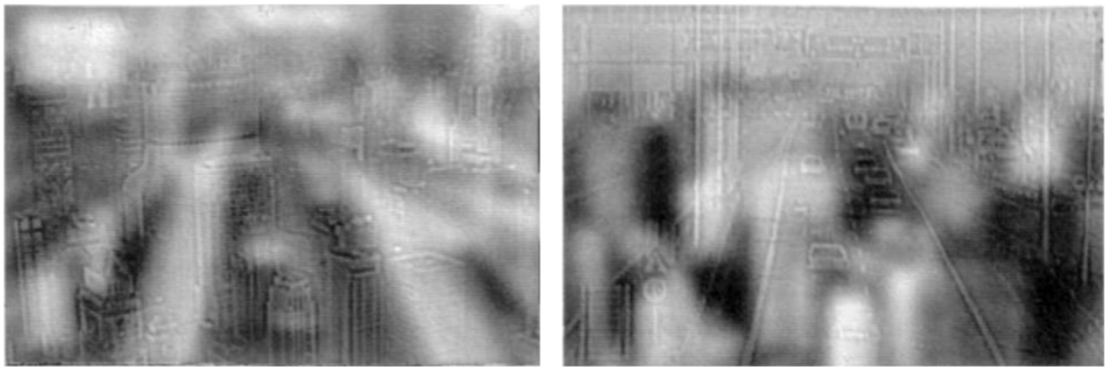

Figure 1.3 – Exemple d'images hybrides utilisées dans l'étude de Schyns et Oliva (1994). À gauche, une image composée des HFS d'une scène de ville et des BFS d'une scène d'autoroute. À droite, l'inverse, une image composée des BFS d'une scène de ville et des HFS d'une scène d'autoroute.

1999 ; Wiesmann et al., 2021). Plus précisément, l'analyse *coarse-to-fine* pourrait faire office de traitement par défaut, qui laisserait la place à un traitement *fine-to-coarse* lorsque les HFS sont pertinentes à la tâche de l'observateur. Nous reviendrons plus précisément sur ce point dans la suite de ce chapitre, en se plaçant dans le contexte du décodage des expressions faciales. Pour finir, certaines études font aussi état d'une latéralisation dans le traitement des fréquences spatiales, dans le sens d'une spécialisation de l'hémisphère droit dans le traitement des informations globales et de l'hémisphère gauche dans le traitement des informations locales (Musel et al., 2013 ; Peyrin et al., 2004).

1.2 La perception visuelle : une analyse fréquentielle - Points clés

- L'information visuelle réceptionnée par les cellules de la rétine est principalement relayée au cortex visuel par la voie rétino-géniculo-striée.
- L'information en BFS est transmise par la voie magnocellulaire, tandis que l'information en HFS est transmise par la voie parvocellulaire. La voie magnocellulaire transmet l'information plus rapidement que la voie parvocellulaire.
- Le traitement de l'information visuelle suivrait par défaut une analyse *coarse-to-fine*, qui pourrait néanmoins être modulée en fonction des besoins de la tâche.

1.3 Attention visuelle et mouvements oculaires

Dans la vie quotidienne, nous faisons face à une image complète du monde visuel, qui rassemble une grande quantité d'informations. Bien que nous puissions avoir l'impression de percevoir de manière stable et uniforme le monde qui nous entoure, notre capacité à traiter l'information disponible est limitée. En effet, les informations visuelles qui parviennent à notre cerveau sont loin d'être complètes, mais notre système visuel est capable d'optimiser

le traitement de l'information. Cela passe notamment par une sélection de l'information pertinente, parmi l'information non pertinente. L'attention visuelle est au coeur de ce processus de sélection. Elle nous permet de traiter sélectivement la grande quantité d'informations, en privilégiant certains aspects ou endroits de la scène et en en ignorant d'autres. L'attention visuelle est très liée aux mouvements oculaires, car, la plupart du temps, la cible d'un mouvement oculaire correspond à la zone sur laquelle on va focaliser notre attention. Cette section a pour but de présenter certains mécanismes impliqués dans l'allocation de l'attention visuelle et la programmation de mouvements oculaires.

1.3.1 L'origine de l'étude des mouvements oculaires comme reflet de l'attention visuelle

L'attention visuelle peut être allouée à un endroit particulier de la scène avec ou sans mouvements oculaires (Posner, 1980). Lorsqu'elle est déplacée sans mouvements oculaires, on parle d'attention *covert* (cachée). Sinon, lorsqu'elle est déplacée vers un endroit de la scène avec des mouvements oculaires, on parle d'attention *overt* (manifeste). L'attention *covert* et l'attention *overt* ne sont pas nécessairement indépendantes. Par exemple, selon la théorie de l'attention prémotrice, l'attention visuelle pourrait être une conséquence de la planification d'un mouvement oculaire (Rizzolatti et al., 1987). Cette idée est appuyée par des données qui montrent que l'allocation de l'attention et la programmation de saccades impliquent des réseaux neuraux communs (Corbetta et al., 1998). Certaines études ont aussi montré que la modification des caractéristiques motrices d'une saccade entraîne des changements proportionnels dans la distribution de l'attention visuelle (Collins et Doré-Mazars, 2006 ; Collins et al., 2010). Ainsi la préparation motrice pourrait contribuer, au moins en partie, à l'allocation de l'attention visuelle.

Dans cette thèse, nous nous intéressons plus particulièrement à l'attention *overt*, au travers des mécanismes impliqués dans la programmation d'un mouvement oculaire. Nous pouvons dissocier trois types de mouvements oculaires : les saccades, les mouvements de poursuite et les mouvements fixationnels. Les saccades sont les mouvements rapides des yeux qui permettent de mettre une région d'intérêt au centre de la fovéa. Les mouvements de poursuite sont les mouvements des yeux qui permettent de suivre un objet qui se déplace. La fin d'une saccade ou d'un mouvement de poursuite correspond au début d'une fixation, c'est-à-dire d'une phase de stabilisation des yeux lors de laquelle l'information visuelle est analysée. Néanmoins, même lors des phases de fixations, nos yeux ne cessent de bouger, sans que nous en ayons conscience. Ces mouvements de faible amplitude ne sont pas consciemment perceptibles, et sont ce que l'on appelle mouvements fixationnels (souvent, trois types de mouvements fixationnels sont distingués : les microsaccades, les tremblements et les dérives ; Martinez-Conde et al., 2004). Dans ce manuscrit, ce sont les saccades oculaires qui nous intéresseront particulièrement.

La recherche scientifique sur les mouvements oculaires a débuté à la fin du 19e siècle, lorsque des méthodes fiables de mesure de la position des yeux ont été mises au point (Buswell, 1935 ; Yarbus, 1967). Notamment, Yarbus fut l'un des premiers à étudier les mouvements oculaires. Dans un livre maintenant classique, il aborde un large éventail de

questions sur la façon dont nous inspectons et percevons des objets complexes (Yarbus, 1967). Par exemple, en enregistrant les mouvements oculaires d'observateurs face à des peintures il a montré que la direction du regard n'est pas aléatoire, mais dépend largement de la tâche de l'observateur. Ce phénomène est illustré sur la Figure 1.4, extraite de l'article de Rolfs (2015). En considérant le tableau *Morning in the Pine Forest*, si la tâche est de trouver l'ours à la fourrure la plus claire, les mouvements oculaires balayeront l'image en faisant des allers-retours entre les ours. En revanche, en l'absence d'une tâche explicite, l'itinéraire des yeux pourra être plus diffus.

Depuis, de nombreuses études se sont intéressées aux facteurs qui influencent la programmation d'un mouvement oculaire. Ces facteurs peuvent, pour la plupart, être classés selon deux types : les facteurs liés à des mécanismes de traitement *bottom-up* (ascendants) et les facteurs liés à des mécanismes de traitement *top-down* (descendants). Les facteurs *bottom-up* sont basés uniquement sur l'information présentée dans le champ visuel, et ils sont donc indépendants de l'observateur. Ils incluent des facteurs comme la couleur, le contraste de luminance ou le mouvement. Les facteurs *top-down*, sont, au contraire, basés sur les connaissances à priori de l'observateur. Ils incluent des facteurs comme la pertinence d'un objet par rapport à la tâche de l'observateur (considérée comme l'un des facteurs les plus déterminants dans l'allocation de l'attention ; Ballard et Hayhoe, 2009 ; Zelinsky et al., 2006), le contenu sémantique ou émotionnel des objets (Hwang et al., 2011 ; Mulckhuyse et al., 2013) ou encore les attentes de l'observateur (Brockmole et Henderson, 2006). Tous ces facteurs vont être intégrés dans les mécanismes de programmation des mouvements oculaires, notamment par l'intermédiaire de ce que l'on appelle des cartes de saillance et des cartes de priorité. L'information encodée dans ces cartes évolue au cours du temps avec l'intégration de l'information visuelle. Si les BFS sont d'abord encodées, comme le suppose le modèle *coarse-to-fine*, alors l'information issue des BFS arrivera en premier aux cartes.

1.3.2 Des cartes de saillance pour l'intégration des facteurs *bottom-up*

La notion de carte de saillance est centrale aux théories de l'attention visuelle. Elle peut s'apparenter à la notion de carte maîtresse, introduite par Treisman et Gelade (1980) dans le cadre de la théorie de l'intégration de caractéristiques. Nous pouvons définir une carte de saillance comme un espace en deux dimensions, à l'image de la scène visuelle, dans lequel chaque localisation de la scène possède un poids. Ce poids correspond à une valeur de saillance, qui dépend des caractéristiques physiques des objets par rapport à leur contexte. Plus une localisation de la scène est saillante, plus elle pourra attirer l'attention. Il existe de nombreux attributs physiques différents qui peuvent rendre un objet plus saillant, comme sa couleur, son orientation, son intensité, ou son mouvement. Ces caractéristiques seraient extraites en parallèle, et représentées dans des cartes corticales topographiques distinctes, appelées cartes d'attributs, avant d'être fusionnées dans la carte de saillance. Plus un objet sera saillant, plus il va se voir attribuer un poids important sur la carte. Ce poids se traduit au niveau neural par des activations plus fortes à la

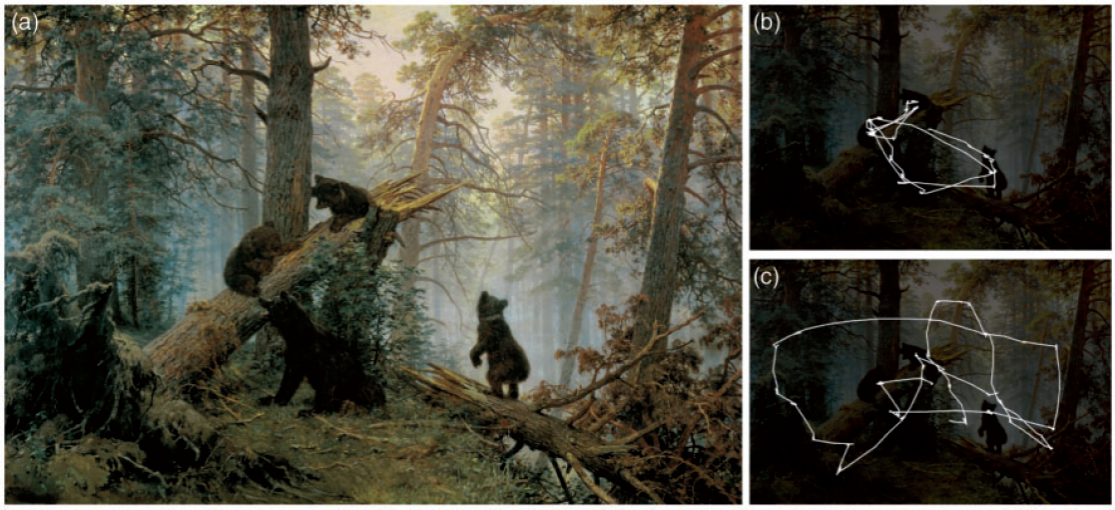

Figure 1.4 – Mouvements oculaires en fonction de la tâche, exemple avec le tableau *Morning in the Pine Forest* (1889) des artistes russes Ivan Shishkin et Konstantin Savitsky. **(a)** Stimulus original. **(b)** Parcours de l'observateur pour trouver l'ours à la fourrure la plus claire. **(c)** Parcours de l'observateur pour une exploration libre de la scène. Figure extraite de Rolfs (2015).

localisation de l'objet sur la carte ; les cellules correspondant à des cartes de saillance ayant une organisation rétinotopique. La carte de saillance code ainsi l'information issue des facteurs *bottom-up*. En 1998, Itti et Koch ont proposé par l'intermédiaire des cartes de saillance l'un des premiers modèles de l'attention visuelle *bottom-up*. Depuis, de nombreux modèles computationnels ont été proposés pour calculer des cartes de saillance à partir d'images, et faire des prédictions concernant les zones qui vont être les plus regardées par des observateurs en exploration libre (pour une revue de la littérature, voir Borji et Itti, 2013, ou, plus récemment, Bylinskii et al., 2016 ; Krasovskaya et MacInnes, 2019).

Au niveau cérébral, plusieurs régions correspondent au profil des cartes de saillance. En fait, il semble qu'il n'y a pas une carte unique et commune dans le cerveau, mais qu'un certain nombre d'aires cérébrales, corticales et sous-corticales, travaillent ensemble pour déterminer la cible d'un mouvement oculaire. Des cartes de saillance seraient présentes au sein du cortex visuel primaire et des couches superficielles du colliculus supérieur (CS ; Klink et al., 2014 ; Theeuwes, 2019 ; Veale et al., 2017 ; White, Kan et al., 2017 ; Zhang et al., 2012 ; Zhaoping, 2002).

1.3.3 Des cartes de priorité pour l'intégration des facteurs *bottom-up* et *top-down*

Comme nous l'avons évoqué précédemment, les facteurs *bottom-up*, représentés sur la carte de saillance, ne suffisent pas à expliquer l'allocation de l'attention visuelle. Pour modéliser l'intégration de l'information liée à la fois aux facteurs *bottom-up* et aux facteurs *top-down*, la littérature semble converger en faveur de l'existence d'une carte de priorité (pour des revues de la littérature, voir Belopolsky, 2015 ; Bisley et Mirpour, 2019 ; Fecteau et Munoz, 2006 ; Klink et al., 2014 ; Zelinsky et Bisley, 2015). Similairement à une carte de saillance, une carte de priorité est un espace en deux dimensions, à l'image de la

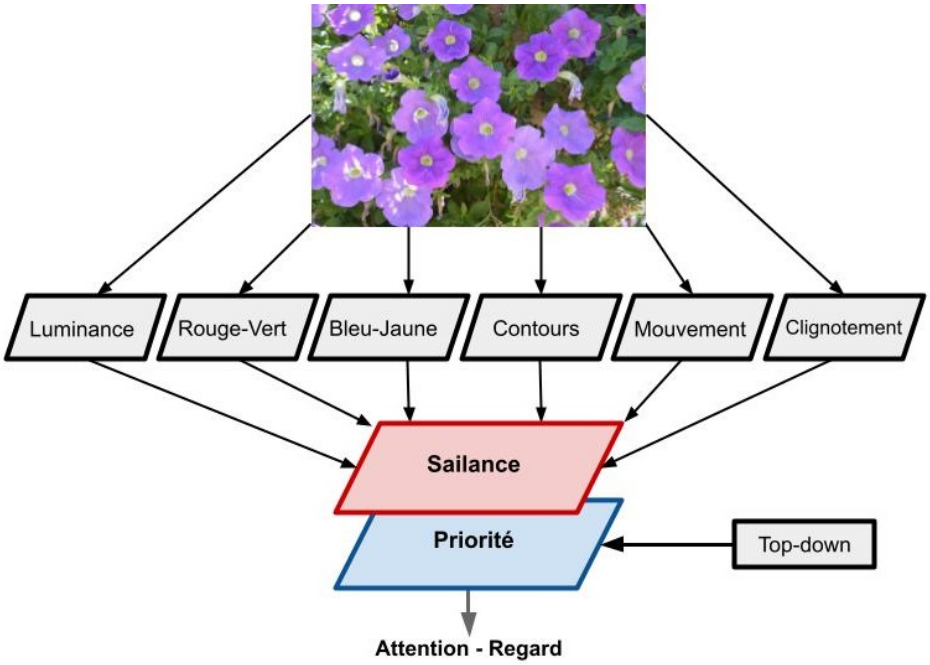

Figure 1.5 – Représentation visuelle des cartes d'attributs, de saillance et de priorité. L'entrée visuelle est décomposée selon différents attributs (gris), qui sont intégrés pour former une carte de saillance (violet). Une carte de priorité (bleu) combine les entrées de la carte de saillance avec des signaux *top-down*, dépendants de l'observateur, pour déterminer la cible du prochain mouvement oculaire. Figure adaptée de White, Kan et al. (2017).

scène visuelle, dans lequel chaque localisation de la scène possède un poids. Ce poids correspond à une valeur de priorité qui est influencée à la fois par des facteurs *bottom-up* et des facteurs *top-down*. L'information sur la saillance *bottom-up* est transmise à la carte de priorité depuis les cartes de saillance. Les interactions entre les différentes cartes (c'est-à-dire entre les cartes d'attributs, de saillance et de priorité) sont illustrées sur la Figure 1.5. Ensuite, la Figure 1.6 présente une carte de priorité hypothétique pour une scène visuelle donnée. Dans cet exemple, on suppose que le but de l'observateur est de trouver le cercle parmi les triangles. Chaque objet de la scène est associé à un poids sur la carte. Le triangle jaune est représenté par un poids légèrement plus élevé que les autres triangles, car il a une couleur qui le distingue des autres triangles. Le cercle est représenté par un poids beaucoup plus élevé, car, en plus d'avoir une forme qui le distingue des autres objets, il est pertinent pour la tâche.

Au niveau cérébral, des cartes de priorité seraient présentes au sein du cortex interpariétal latéral, du champ visuel frontal et des couches intermédiaires du CS. Dans une revue de la littérature sur le codage neural des cartes de priorité, Bisley et Mirpour (2019) suggèrent que le cortex interpariétal latéral représente la priorité attentionnelle, qui change après chaque saccade. Le champ visuel frontal recevrait l'information issue du cortex interpariétal latéral, mais pourrait choisir de programmer un mouvement oculaire ou non. Le but d'une saccade serait finalement représenté dans les couches intermédiaires du CS. Ainsi, contrairement au champ visuel frontal, le CS n'est actif que s'il y a effectivement le

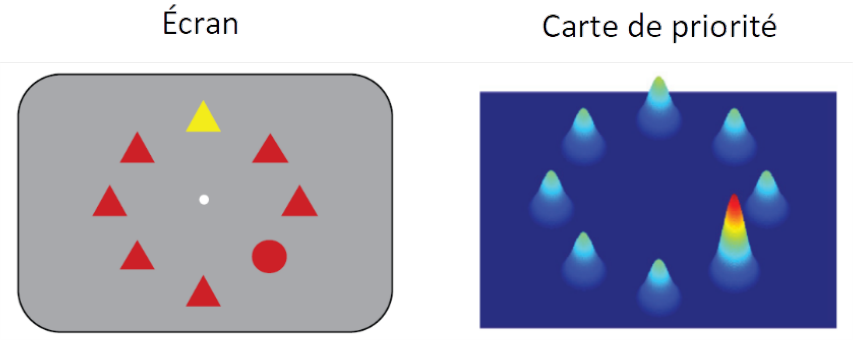

Figure 1.6 – Carte de priorité hypothétique (droite) en réponse à un ensemble de stimuli (gauche), dans lequel le sujet doit trouver un cercle. Chaque stimulus est associé à un poids sur la carte. Le triangle jaune est représenté par un poids légèrement plus élevé que les autres triangles, car il est plus saillant (sa couleur le distinguant des autres triangles). Le cercle est représenté par une réponse beaucoup plus élevée car, en plus d'être saillant (sa forme le distinguant des autres objets), il est pertinent pour la tâche. Figure extraite de Bisley et Mirpour (2019).

déclenchement d'un mouvement oculaire. Le CS va ensuite envoyer un signal au tronc cérébral pour déclencher une réponse motrice. Ces cartes de priorité pourraient recevoir des informations sur la saillance depuis V1 et depuis les couches superficielles du CS. Les informations *top-down* sont plus susceptibles de provenir de régions telles que le cortex cingulaire antérieur et le cortex orbitofrontal (Kennerley et al., 2011 ; Klink et al., 2014). Les informations émotionnelles seraient transmises par l'amygdale, un point sur lequel nous reviendrons plus tard dans ce chapitre.

1.3.4 La programmation de saccades dans les cartes de priorité

Ainsi, la programmation de saccades oculaires aurait lieu dans une carte de priorité telle que définie précédemment. Dans cette section, nous allons préciser les mécanismes qui entrent en jeu dans la phase de programmation d'une saccade, ainsi que leur impact sur les points d'arrivée des saccades.

1.3.4.1 Une programmation en parallèle ?

Même si les saccades oculaires sont effectuées de manière sérielle, un certain nombre d'études suggère que le système oculomoteur est capable de développer plusieurs programmes saccadiques en même temps (Dorris et al., 2007 ; Findlay et Walker, 1999 ; McPeek et al., 2000 ; Trappenberg et al., 2001 ; Walker et McSorley, 2006). Par exemple, McPeek et al. (2000) ont fait une expérience dans laquelle les participants devaient faire une saccade vers une cible (un losange de couleur) en présence de distracteurs (deux losanges d'une autre couleur). Lorsqu'il y avait une forte compétition entre la cible et le distracteur (par exemple lorsque le distracteur était de la couleur des cibles précédentes), les participants étaient plus susceptibles de faire une saccade vers le distracteur. De telles saccades, dites erreurs, étaient souvent suivies, après un intervalle inter-saccadique très

court (entre 10 et 100 ms), d'une seconde saccade, dirigée vers la cible. La brièveté de ces intervalles inter-saccadiques suggère que la programmation des deux saccades (une vers le distracteur et une vers la cible) se chevauche dans le temps ; l'intervalle inter-saccadique étant en effet trop court pour que la scène puisse être analysée, et qu'il y ait la préparation de la prochaine saccade. Ainsi, les différents programmes saccadiques pourraient se développer en parallèle au niveau de la carte de priorité du CS. De plus, les programmes dont ne résulte pas le déclenchement d'une saccade pourraient interférer avec la saccade en cours, ce qui expliquerait certains effets comportementaux (Coe et al., 2019 ; Trappenberg et al., 2001). Par exemple, les résultats observés dans des tâches d'anti-saccades, dans lesquelles les participants doivent faire une saccade dans la direction opposée à un stimulus apparaissant dans leur champ de vision. Les interférences causées par une saccade qui se développerait de manière réflexive vers le stimulus viendraient ralentir le développement de la saccade vers la direction opposée. Ceci expliquerait le temps additionnel requis pour générer une anti-saccade, en comparaison à une pro-saccade (dirigée vers le stimulus).

1.3.4.2 Les points d'arrivée des saccades comme reflet de la compétition entre différentes localisations

Ces interférences, causées par les localisations qui ont un poids fort dans la carte de priorité lors de la programmation de saccades, vont aussi se refléter sur la position du point d'arrivée d'une saccade. En effet, les points d'arrivée des saccades sont le résultat de la compétition entre plusieurs localisations, et pourraient refléter la distribution de l'attention pendant la programmation des saccades. Ainsi, lorsqu'une saccade doit être exécutée vers une cible (par exemple un anneau gris) en présence d'un distracteur proche (par exemple un anneau noir), la saccade atterrit quelque part entre la cible et le distracteur. Ce phénomène est souvent appelé effet global ou *saccade averaging* (Findlay, 1982 ; Van der Stigchel et Nijboer, 2013). Plusieurs auteurs suggèrent que, dans ce cadre, l'attention est dirigée vers l'emplacement de la cible et du distracteur, plutôt que vers l'emplacement exact où la saccade atterrit (qui se trouve entre la cible et le distracteur ; Van der Stigchel et de Vries, 2015 ; Wollenberg et al., 2018). Par exemple, Van der Stigchel et al. (2015) ont étudié le couplage attentionnel entre les points d'arrivée des saccades et l'allocation de l'attention pendant la préparation de la saccade. Ils ont présenté 10 cercles, 5 à gauche et 5 à droite de l'écran, à des participants qui devaient faire une saccade vers une cible. La cible correspondait au remplissage de l'un des cercles par du gris. En même temps, un distracteur pouvait apparaître sous la forme d'un remplissage en noir d'un autre cercle. Avant le déclenchement de la saccade, une tâche de discrimination de l'orientation des barres à l'intérieur des cercles était utilisée afin d'évaluer l'allocation de l'attention. Les résultats ont montré que la discrimination était meilleure à l'emplacement de la cible et du distracteur par rapport à l'emplacement intermédiaire (là où atterrit, le plus souvent, la saccade). Les performances étaient similaires pour la cible et le distracteur, ce qui suggère que l'attention n'est pas strictement attachée au but du participant, et que le point d'arrivée n'atterrit pas nécessairement sur la zone qui attire le plus l'attention. Cette hypothèse est cohérente avec les résultats d'enregistrements neurophysiologiques dans des

configurations similaires, qui ont montré que les pics d'activité dans le CS correspondent aux cellules codant les emplacements de la cible et du distracteur (Edelman et Keller, 1998).

1.3 Attention visuelle et mouvements oculaires - Points clés

- L'allocation de l'attention fait intervenir des facteurs *bottom-up*, liés aux caractéristiques physiques des stimuli, et des facteurs *top-down*, liés aux caractéristiques de l'observateur.
- Les cartes de saillance permettent de modéliser l'intégration des facteurs *bottom-up*, tandis que les cartes de priorité permettent de modéliser l'intégration des facteurs *bottom-up* et *top-down*.
- Dans le cerveau, la phase finale de la programmation des saccades se fait dans une carte de priorité localisée au niveau des couches intermédiaires du CS.
- Les points d'arrivée des saccades peuvent refléter les interactions qui ont lieu dans cette carte de priorité.

1.4 Les visages : des stimuli particuliers

Naturellement, certains stimuli semblent attirer l'attention plus que d'autres. C'est le cas des visages, qui nous intéressent particulièrement dans ce travail de thèse au travers de leur capacité à exprimer des émotions. Les visages fournissent une multitude d'informations essentielles à la survie, au bien-être émotionnel, ou encore à la fonction sociale des individus. Ainsi, il n'est pas étonnant de voir qu'ils bénéficient d'un traitement privilégié au sein du système visuel. Les localisations de la scène qui contiennent un visage pourraient ainsi systématiquement bénéficier d'un poids important dans la carte de priorité. Dans cette partie, nous nous intéresserons à une sélection d'études comportementales qui mettent en évidence une capture de l'attention par les visages. Nous nous interrogerons également sur les mécanismes impliqués, en lien avec le traitement des fréquences spatiales.

1.4.1 Une capture attentionnelle

Le statut particulier des visages dans les mécanismes d'allocation de l'attention visuelle a été mis en évidence dès les premiers enregistrements des mouvements oculaires. Ainsi, les travaux de Yarbus en 1967 montraient déjà que, lors de l'exploration visuelle de peintures, sans consigne particulière et en présence de visages, ceux-ci faisaient l'objet de nombreuses fixations. Depuis, de nombreuses études ont montré, dans des paradigmes d'exploration libre, que les visages ont tendance à être fixés très rapidement, et pendant une grande partie du temps d'exploration (voir par exemple Cerf et al., 2009 ; Coutrot et Guyader, 2014 ; Marat et al., 2013). Aussi, d'autres paradigmes ont permis de mettre en évidence le fait que les visages capturent l'attention plus facilement que d'autres objets. Par exemple, les paradigmes de recherche visuelle, dans lesquels des grilles composées d'images de différentes catégories sont présentées à des observateurs, qui doivent trouver

une catégorie cible. Plusieurs études ont montré que, lorsque la cible est un visage, les participants sont plus efficaces (c'est-à-dire plus rapides pour trouver la cible) que si la cible est, par exemple, un véhicule ou une maison (Hershler et Hochstein, 2005, 2006 ; VanRullen, 2006). On parle d'effet *pop-out* des visages, qui souligne leur accessibilité et leur capacité à se démarquer rapidement des autres objets. La Figure 1.7 présente une grille d'images telle que celles utilisées dans les études qui montrent un tel effet.

Il semble aussi que les visages soient capables d'attirer l'attention même lorsque les observateurs doivent effectuer une tâche précise, dans laquelle ils ne sont pas pertinents. Par exemple, Sato et Kawahara (2015) ont demandé à des participants d'identifier une lettre cible parmi des lettres qui défilaient rapidement à l'écran, en ignorant un distracteur qui pouvait apparaître en périphérie de la cible, ou de la lettre précédant la cible. Ils ont montré que, si le distracteur est un visage, les performances de détection de la cible sont altérées, ce qui n'est pas le cas si le distracteur est un simple rectangle ou un visage *scramble* (c'est-à-dire une image de visage dont la structure est altérée, mais dont le contenu fréquentiel et chromatique est intact). Une autre étude a répliqué ce résultat, en montrant néanmoins qu'il était seulement présent quand le temps séparant le distracteur et la cible était faible (<100 ms ; Ariga et Arihara, 2018). Les auteurs en ont conclu que les visages peuvent attirer l'attention rapidement, et ce même si cela va à l'encontre des objectifs de l'observateur, une hypothèse également appuyée par des études en oculométrie (Devue et al., 2012 ; Devue et Grimshaw, 2017). Par exemple, dans une étude de Devue et al. (2012), un cercle de points était affiché à l'écran, et les participants devaient effectuer une saccade vers un point d'une certaine couleur. Des images de différents objets étaient affichées dans un cercle concentrique, à l'intérieur du cercle de points. Les résultats ont montré que, lorsqu'ils n'étaient pas associés au point cible, les visages attiraient plus souvent la première saccade que les autres objets. De plus, ils favorisaient la détection de la cible lorsqu'ils étaient proches d'elle.

1.4.2 Une détection rapide : intérêt du paradigme de choix saccadique

Bien que nous n'en présentons pas dans ce chapitre une liste exhaustive, cette capacité des visages à attirer l'attention a été mise en évidence dans de nombreuses tâches. La plupart ne permettant pas d'identifier précisément le décours temporel de la détection des visages. Les mouvements oculaires sont particulièrement bien adaptés à la mesure des vitesses de traitement, car certains d'entre eux peuvent être initiés très vite (Fischer et Weber, 1993). En 2006, Kirchner et Thorpe ont présenté une étude dans laquelle les mouvements oculaires étaient utilisés comme réponse pour déterminer la vitesse de traitement de scènes naturelles. Plus précisément, deux images de catégories différentes étaient affichées simultanément à gauche et à droite du centre d'un écran. L'une des images contenait un animal, tandis que l'autre contenait une scène naturelle sans animal. Les participants devaient faire une saccade le plus rapidement possible vers la scène qui contenait un animal. Ainsi, la saccade était utilisée comme une réponse comportementale. Les saccades étant effectuées plus rapidement que les réponses manuelles,

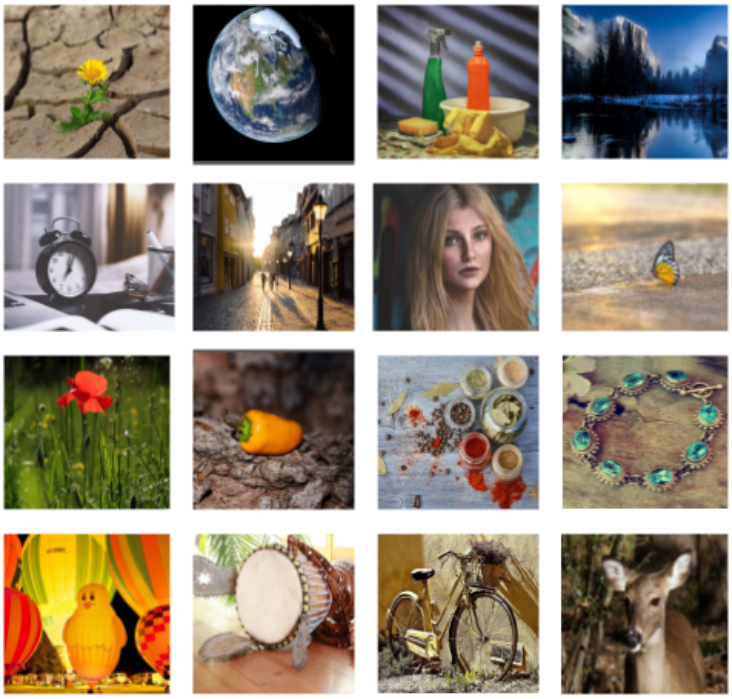

Figure 1.7 – Grille d'images avec un visage et une variété de distracteurs. Cette grille est similaire à celles qui sont utilisées dans des paradigmes de recherche visuelle, qui ont montré un effet *pop-out* des visages (c'est-à-dire des temps de recherche plus faibles lorsque la cible est un visage ; voir par exemple Hershler et Hochstein, 2005, 2006).

cela permet d'évaluer plus précisément la vitesse de catégorisation des scènes. Au final, leur étude a mis en évidence un traitement rapide des scènes, avec la détection d'un animal en seulement 120 ms.

Ce paradigme est connu sous le nom de paradigme de choix saccadique, et il a depuis été repris dans de nombreuses études (Bannerman, Milders et Sahraie, 2009 ; Boucart et al., 2016 ; Calvo et Nummenmaa, 2011 ; de Lissa et al., 2021 ; D'Hondt et al., 2016 ; Guyader et al., 2017 ; Honey et al., 2008 ; Kauffmann et al., 2021 ; Kauffmann et al., 2019). Notamment, il a permis de mettre en évidence une détection très rapide des visages, en comparaison à d'autres stimuli, comme les véhicules, les bâtiments ou les animaux (Crouzet et al., 2010 ; Guyader et al., 2017 ; Kauffmann et al., 2019). Ainsi, Crouzet et al. (2010) ont repris ce paradigme en opposant systématiquement des visages et des véhicules (Expérience 2). Leurs résultats ont montré que les latences des saccades (c'est-à-dire les temps qui séparent l'apparition des images et le déclenchement de la saccade) sont, en moyenne, plus faibles quand la cible est le visage que le véhicule, et les participants font moins d'erreurs (c'est-à-dire moins de saccades en direction du distracteur). Les auteurs ont aussi calculé le temps minimal dont les participants avaient besoin pour détecter la cible : 100 ms pour les visages, et 150 ms pour les véhicules. Ces valeurs ont été calculées à partir de la distribution des latences des saccades correctes et des saccades erreurs, représentée sur la Figure 1.8. Plus précisément, les auteurs ont divisé les essais de tous les

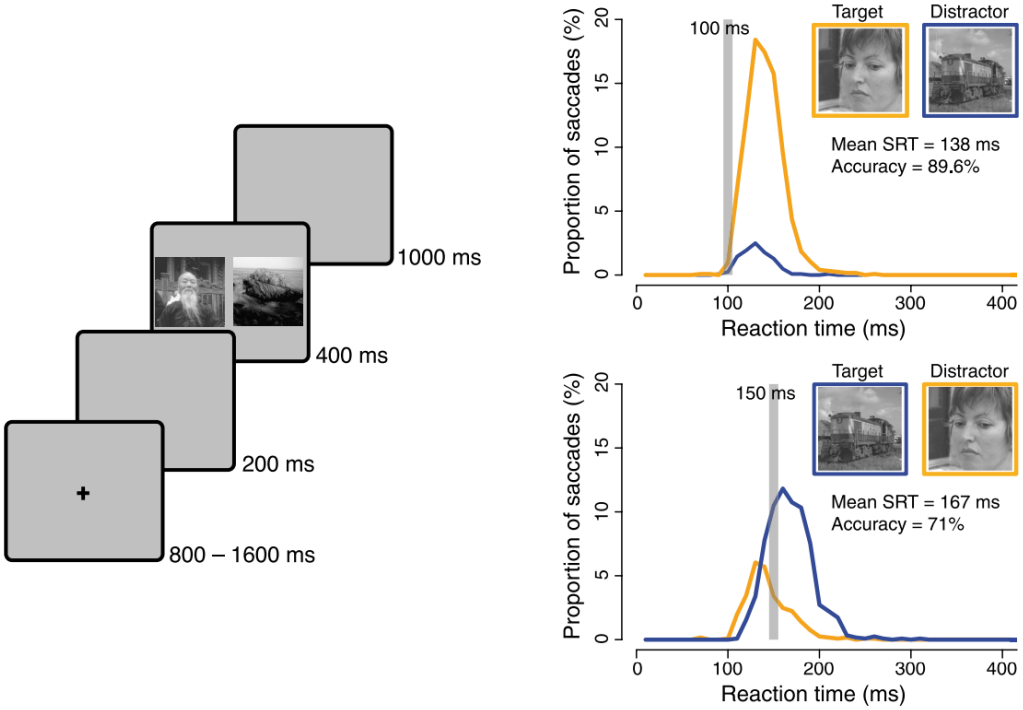

Figure 1.8 – Illustration d'un essai en paradigme de choix saccadique (gauche). Les observateurs doivent fixer une croix au centre de l'écran pendant un temps pseudo-aléatoire (800 - 1600 ms). La croix laisse place à un écran gris pendant 200 ms, puis deux images sont affichées à gauche et à droite de l'écran pendant 400 ms. Les observateurs doivent, le plus rapidement possible, faire une saccade vers l'image qui contient la catégorie cible, et ont ensuite 1000 ms pour se préparer à l'essai suivant. Répartition des latences des saccades (droite) lorsque la cible est un visage (haut) ou un véhicule (bas). La barre grise indique la fenêtre temporelle à partir de laquelle les participants sont capables de détecter de manière fiable le visage ou le véhicule. Figures extraites de Crouzet et al. (2010).

participants en intervalles temporels de 10 ms. Si cinq intervalles consécutifs contenaient significativement plus de saccades correctes (c'est-à-dire dirigées vers la cible) que de saccades erreurs (c'est-à-dire dirigées vers le distracteur), le premier intervalle était choisi comme le temps minimal dont les participants ont besoin pour détecter la cible. De manière intéressante, sur la Figure 1.8 nous pouvons voir que, même lorsque la cible est un véhicule, les saccades rapides ont tendance à aller plus souvent vers le visage. Les auteurs ont interprété cet effet comme une preuve que les saccades vers les visages sont déclenchées de manière automatique, c'est-à-dire rapidement et indépendamment du but de l'observateur.

1.4.3 Un traitement holistique

Une telle capture de l'attention par les visages nous amène à nous questionner sur les mécanismes impliqués. Souvent, deux hypothèses sont distinguées pour caractériser le traitement des visages : celle d'un traitement holistique/configural et celle d'un traitement analytique. Le traitement analytique suppose que la perception d'un objet est basée sur l'analyse des différentes parties qui le composent. Ainsi, dans le cadre de la perception des

visages, c'est l'analyse indépendante de zones comme les yeux, le nez, ou la bouche qui en serait la base. Au contraire, le traitement holistique/configural suppose l'intégration de l'objet comme un tout, plutôt que comme un ensemble de caractéristiques indépendantes [4]. Ces concepts de traitement analytique et holistique sont en fait communs à différents domaines de la perception, mais ont été étudiés de manière extensive dans le domaine de la perception des visages.

S'il est admis que la plupart des objets sont traités de manière analytique, la perception des visages se ferait le plus souvent sur la base d'un traitement holistique (Gold et al., 2012 ; Piepers et Robbins, 2012 ; Richler et Gauthier, 2014 ; Rossion, 2013 ; Rossion et Retter, 2020). Ce traitement peut, par exemple, être mis en évidence dans un paradigme *part/whole*, introduit par Tanaka et Farah en 1993. Dans leur étude, les participants devaient d'abord étudier un visage entier. Ensuite, leur capacité à reconnaître une partie du visage était testée, soit quand cette partie était présentée seule, soit quand elle était présentée dans le contexte du visage entier. L'hypothèse des auteurs était que, si une partie du visage est intégrée dans la mémoire comme une représentation du visage entier (perception holistique), la reconnaissance de cette partie devrait être plus élevée lorsqu'elle est présentée dans le visage entier plutôt que seule. Leurs résultats ont mis en évidence un tel effet, qui a depuis été largement répliqué (pour une revue des apports théoriques de ce paradigme dans la recherche sur le traitement des visages voir Tanaka et Simonyi, 2016). De manière similaire, plusieurs études ont mis en évidence le fait que deux moitiés supérieures identiques d'un visage sont perçues comme différentes lorsque leurs moitiés inférieures appartiennent à des visages différents. On parle alors d'un effet d'illusion composite, qui souligne une dépendance entre les parties du visage, et donc un traitement holistique (Murphy et al., 2017 ; Rossion, 2013 ; Young et al., 2013). De plus, ce traitement holistique pourrait expliquer l'effet d'inversion des visages, qui caractérise une diminution des performances dans des tâches impliquant des visages lorsque ceux-ci sont présentés à l'envers plutôt qu'à l'endroit (Valentine, 1988 ; Yin, 1969). En effet, l'inversion des visages pourrait empêcher le traitement holistique des visages en faveur d'un traitement plus local (Farah et al., 1995 ; Rossion, 2009).

Pour finir, plusieurs auteurs suggèrent que le traitement holistique des visages intervient à un stade précoce de la perception des visages (Rossion, 2009 ; Taubert et al., 2011 ; Tsao et Livingstone, 2008 ; pour un point de vue différent, voir Maurer et al., 2002). Ce traitement serait basé sur un modèle grossier de la représentation des visages, et permettrait de les distinguer des autres objets.

4. Le traitement holistique et le traitement configural ont parfois été considérés comme identiques et ont parfois été distingués. Dans ce chapitre, nous considérons qu'ils ne diffèrent pas, mais la définition proposée correspond plutôt à celle d'un traitement holistique. Le traitement configural suppose une analyse des relations entre les parties, mais pas nécessairement de toutes les parties en même temps, contrairement au traitement holistique. Pour une revue de la littérature sur l'utilisation de ces termes dans le cadre de la perception des visages, voir Piepers et Robbins (2012).

1.4.4 Rôle des basses fréquences spatiales

En supposant que la perception des visages soit basée sur un tel modèle grossier, nous pouvons naturellement faire le parallèle avec les fréquences spatiales. En ce sens, une étude menée par Goffaux et al. (2005) a directement mis en évidence l'importance des BFS dans le traitement holistique des visages, et des HFS dans leur traitement analytique. Dans cette étude, trois visages étaient présentés à des participants qui devaient associer deux visages identiques, l'un des visages étant différent au niveau de sa configuration (par l'intermédiaire d'une manipulation des relations spatiales au niveau des yeux) ou de ses parties (par l'intermédiaire d'un remplacement des yeux par ceux d'un autre visage). Lorsque la différence se situait au niveau de la configuration, les performances étaient meilleures en BFS qu'en HFS, alors que l'inverse était observé lorsque la différence se situait au niveau des parties. Ainsi, ce serait les BFS qui supporteraient le traitement holistique des visages. Cette hypothèse a également été mise en évidence par Goffaux et Rossion (2006) en utilisant le paradigme *part/whole* introduit dans la section précédente. Les auteurs ont répliqué l'effet classique de l'avantage des visages entiers, et ont montré qu'il était deux fois plus important quand les images étaient présentées en BFS plutôt qu'en HFS. Les auteurs font également un lien entre le traitement holistique et la voie magnocellulaire, ainsi qu'entre le traitement analytique et la voie parvocellulaire, ce qui induirait une précédence du traitement holistique par rapport au traitement analytique.

En lien avec l'idée d'un traitement holistique précoce, basé sur les BFS, d'autres études ont mis en avant l'importance de l'information globale plutôt que locale dans la détection des visages (Awasthi et al., 2011 ; Goffaux et al., 2005 ; Goffaux et al., 2011 ; Guyader et al., 2017 ; Peters et al., 2018 ; Quek et al., 2018). Par exemple, en utilisant un paradigme de choix saccadique, Guyader et al. (2017) ont observé, en accord avec les résultats présentés dans la section 1.4.2, une meilleure détection des visages par rapport aux véhicules ou aux animaux. Ils ont également observé que le temps minimal dont les participants avaient besoin pour détecter les visages était plus court lorsque les images étaient présentées en BFS qu'en HFS (140 et 160 ms pour les BFS et les HFS, respectivement ; le temps minimal requis lorsque les images étaient non filtrées était de 130 ms), une différence qui n'était pas observée pour les véhicules et les animaux. De plus, en se basant sur une analyse des statistiques des images dans le domaine de Fourier, les auteurs ont souligné la prépondérance du contenu en BFS dans le spectre d'amplitude des visages. Du côté des études en neuroimagerie, une étude utilisant une technique d'Imagerie par Résonance Magnétique fonctionnelle (IRMf), a mis en avant une activité neurale plus élevée au sein des aires sélectives aux visages lorsque des visages étaient présentés en BFS ou non filtrés plutôt qu'en HFS (Goffaux et al., 2011). Ce résultat n'était obtenu qu'avec une présentation rapide des visages (75 ms). Pour une présentation plus longue (150 ms), le schéma était inversé avec une activité plus forte en réponse à des visages présentés en HFS. Au final, même si l'extraction des fréquences spatiales dans les visages dépend de nombreux facteurs (pour une revue de la littérature récente, voir Jeantet et al., 2018), la détection des visages semble être basée sur une extraction précoce de la configuration

spatiale à partir des BFS.

1.4 Les visages : des stimuli particuliers - Points clés

- Les visages attirent particulièrement le regard et l'attention, et cela indépendamment de la volonté de l'observateur.
- Lorsqu'ils sont opposés à une image d'une autre catégorie (par exemple un véhicule) ils peuvent être détectés et déclencher une saccade en seulement 100 ms.
- La rapidité du traitement des visages pourrait s'expliquer par un traitement rapide et grossier de la configuration des visages, basé sur l'extraction des BFS.

1.5 L'expression faciale des émotions : reconnaissance et attributs

Nous avons vu dans la section précédente que les visages sont traités très efficacement par le système visuel. Cette efficacité peut se justifier par la multitude d'informations essentielles que les visages transmettent. Par exemple, l'état émotionnel d'un individu se reflète au travers des expressions faciales, c'est-à-dire au travers de changements morphologiques, comme l'ouverture de la bouche, le plissement des yeux ou du nez. L'expression des émotions est essentielle à des interactions sociales adaptées, que ce soit pour comprendre l'état interne d'un interlocuteur ou communiquer son propre état interne. Elle a aussi un intérêt pour la survie, car elle peut alerter en présence d'une menace. Pour être utiles, les expressions faciales doivent être facilement reconnaissables. C'est pour cela qu'une émotion est généralement associée à une expression dont la configuration spatiale est stéréotypée. Dans cette partie, nous allons définir les expressions faciales et leur traitement en nous intéressant en particulier aux questions suivantes : que sont-elles ? Comment sont-elles reconnues et quelles parties du visage sont les plus importantes ?

1.5.1 Théories influentes dans l'étude des expressions faciales

1.5.1.1 Darwin et l'origine des expressions

L'étude scientifique des visages émotionnels a connu un premier évènement marquant à la fin du 19e siècle, en 1872, avec la publication du livre *The expression of emotion in man and animals* de Darwin. À ce moment, Darwin est déjà connu pour ses travaux sur l'origine des espèces. Même si aujourd'hui ses travaux sur les émotions sont moins connus, ils ont fait l'objet d'un engouement populaire conséquent à la sortie du livre, et ont constitué un apport considérable à l'étude des émotions et leur expression (pour une discussion sur l'impact des recherches de Darwin sur la recherche actuelle sur l'expression des émotions, voir par exemple Barrett, 2011 ; Ekman, 2006 ; Gross et Preston, 2020 ; Hess et Thibault, 2009). Plusieurs idées nouvelles sont développées dans le livre, par exemple

sur l'origine des expressions. Darwin propose que certaines expressions sont des réflexes instinctifs, façonnés par l'évolution, tandis que d'autres sont le résultat d'associations ou d'habitudes apprises. Cette idée selon laquelle certaines expressions sont innées est liée a une autre hypothèse présentée dans le livre, qui suggère que les hommes et les animaux partagent des bases communes concernant l'expression des émotions. En effet, Darwin souligne que certaines expressions sont très similaires chez l'homme et l'animal. Concernant la fonction des émotions, Darwin propose que certaines émotions permettent de réagir efficacement à des menaces, tandis que d'autres ont pour but de communiquer et réguler les interactions sociales. Finalement, Darwin aborde aussi le lien entre état d'esprit et expression. Pour lui, un état d'esprit provoque une décharge musculaire qui va l'exprimer ou servir de décharge d'énergie nerveuse (par exemple pour le rire).

1.5.1.2 Émotions de base et système de codage

Une autre approche, dans la continuité de celle de Darwin, a fortement influencé les recherches sur l'expression des émotions. C'est la théorie d'Ekman, sur les émotions de base (Ekman, 1999). Cette théorie est inspirée de travaux sur la reconnaissance d'émotions effectués auprès d'habitants de Nouvelle-Guinée, ayant eu très peu de contacts avec le monde extérieur (Ekman et Friesen, 1971). Selon cette théorie, il y aurait des émotions de base, qui seraient innées et universellement reconnues. La notion "de base" définie par Ekman implique que les émotions soient définies de manière discrète et qu'elles soient distinguables les unes des autres. Elle implique aussi que les émotions soient le fruit d'une adaptation à l'environnement lors de l'évolution. Pour Ekman, les réactions physiologiques associées aux émotions auraient une fonction de préparation à une réaction adaptée. Ainsi, une sensation de peur serait associée à la redirection du flux sanguin des mains vers les jambes pour préparer l'organisme à fuir (Ekman, 2004). L'expression faciale de la peur, au travers de l'ouverture des yeux, du nez et de la bouche, permettrait d'augmenter l'acquisition sensorielle (Susskind et al., 2008). Ainsi, même s'il conçoit qu'une émotion peut être ressentie en l'absence d'interactions avec d'autres organismes, Ekman propose que sa fonction primaire est de mobiliser l'organisme pour faire face rapidement à des rencontres interpersonnelles importantes. Dans ce cadre, il a référencé des émotions, qui sont en fait une famille d'états similaires, qui rentrent dans sa définition des émotions de base. Ces émotions sont : la joie, la peur, la surprise, la colère, le dégoût et la tristesse (moins souvent cité, le mépris peut aussi être considéré comme une émotion de base ; Ekman et Cordaro, 2011). Un exemple de chacune de ces émotions est présenté sur la Figure 1.9. Les autres émotions sont considérées comme des états affectifs, et non comme des émotions (pour un détail des critères qui différencient les émotions de base des autres états internes, voir Ekman et Cordaro, 2011). Cette théorie des émotions de base est encore d'actualité, même si elle a donné lieu certains débats. Par exemple, certains modèles proposent que les émotions soient définies par des dimensions indépendantes plutôt que des catégories distinctes (voir par exemple Feldman Barrett et Russell, 1998 ; Fontaine et al., 2007). Certaines études remettent également en question le caractère universel des émotions de base (Jack et al., 2009 ; Jack et al., 2012 ; Russell, 1994).

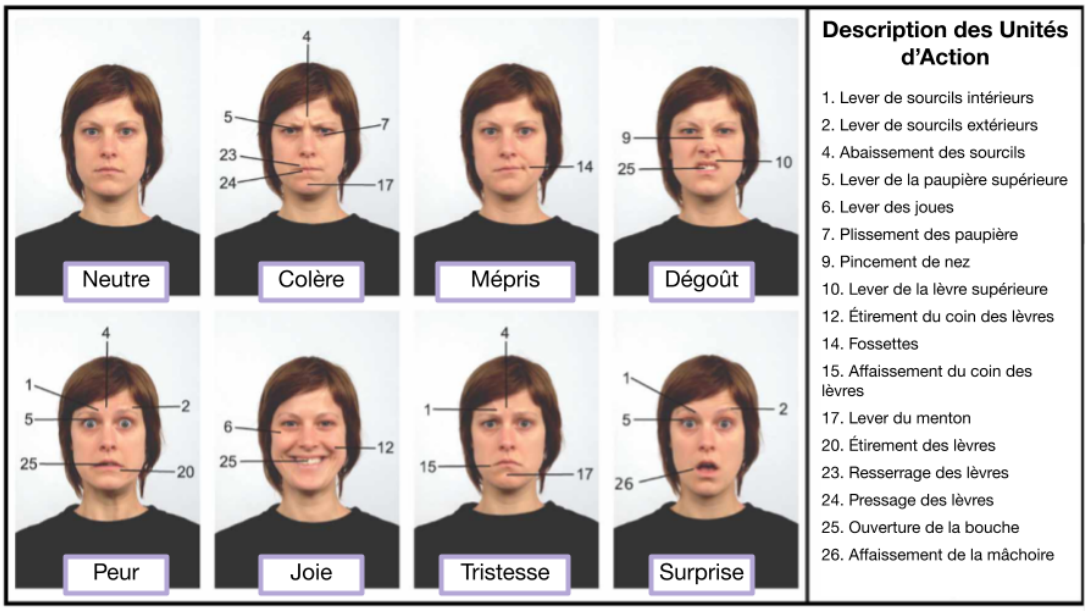

Figure 1.9 – Exemple d'émotions de base selon la classification d'Ekman (2011) et visualisation des unités d'action associées à chacune des émotions. Figure extraite de Langner et al. (2010).

Du fait que les expressions faciales ont une configuration stéréotypée, elles peuvent être décrites par des systèmes de mesures. En ce sens, plusieurs systèmes ont été proposés. Le plus connu est un système de description des expressions faciales nommée le *Facial Action Coding System* (FACS), développé par Ekman et Friesen (1978). Dans ce système, les expressions peuvent être décomposées selon des unités d'action, qui caractérisent des contractions ou des décontractions des différents muscles du visage. Ainsi, ce système est basé sur l'anatomie, et permet de mesurer tous les mouvements faciaux visuellement discernables sur la base de 44 unités d'action et de différents mouvements et positions des yeux et de la tête (pour le manuel, voir Ekman, 2002 ; Ekman et Friesen, 1978 ; pour une revue des utilisations de ce système, voir Rosenberg et Ekman, 2020). Il suppose que la configuration statique, posée et stéréotypée des muscles faciaux fournit des indices suffisants pour reconnaître les émotions. La Figure 1.9 donne un exemple de l'utilisation de ce système de mesure, avec un codage des unités d'action associées à chacune des émotions de base. Par exemple, la joie est associée aux unités d'action 6, 12 et 25, correspondant à un lever des joues, un étirement du coin des lèvres et une ouverture de la bouche, respectivement.

1.5.2 L'expression joyeuse : la mieux reconnue ?

À partir de la définition des émotions de bases, plusieurs études se sont intéressées à notre capacité à reconnaître leur expression faciale. Le plus souvent, dans ces études, il est demandé à des participants de catégoriser une image de visage en fonction de son émotion (par exemple la joie, la peur, la colère, le dégoût, la surprise ou la tristesse). Dans cette section, les résultats de ces études, ainsi que différentes interprétations sont présentés.

1.5.2.1 Résultats comportementaux

Dans la plupart des études sur la reconnaissance des émotions, les performances de catégorisation sont comparées pour six émotions de base : la joie, la peur, la colère, le dégoût, la surprise et la tristesse. Comme le rapporte une revue de la littérature effectuée par Calvo et Nummenmaa en 2016, les scores sont généralement supérieurs à 50% pour toutes les expressions. La plupart du temps, ils sont supérieurs à 70%, sauf parfois pour le dégoût et la peur (voir par exemple Recio et al., 2013) ou pour la tristesse, le dégoût et la peur (voir par exemple Palermo et Coltheart, 2004). Ainsi, certaines expressions semblent mieux reconnues que d'autres. De manière générale, les réponses des participants sont plus rapides et plus précises lorsqu'il s'agit de reconnaître des visages joyeux, puis surpris, en colère et tristes, et elles sont plus longues et moins précises lorsqu'il s'agit de reconnaître des visages apeurés (Blais et al., 2017 ; Calder et al., 2000 ; Calvo et al., 2008 ; Elfenbein et Ambady, 2003 ; Goeleven et al., 2008 ; Langner et al., 2010 ; Palermo et Coltheart, 2004 ; Svard et al., 2012 ; Tottenham et al., 2009 ; Wegrzyn et al., 2017 ; Wilhelm et al., 2014). De plus, certaines expressions sont régulièrement confondues. C'est le cas de la peur et de la surprise ainsi que de la colère et du dégoût (Calvo et Lundqvist, 2008 ; Langner et al., 2010 ; Tottenham et al., 2009). Au final, même si certains facteurs peuvent moduler les performances de reconnaissance, comme l'ouverture de la bouche sur les images (Horstmann et al., 2012 ; Sweeny et al., 2013 ; Tottenham et al., 2009) ou certaines pathologies (Kohler et al., 2010 ; Krause et al., 2021), l'expression joyeuse semble systématiquement mieux reconnue que les autres expressions. De plus, ce résultat se retrouve avec différentes bases de données de photographies. Plusieurs hypothèses ont été proposées pour expliquer cet effet.

1.5.2.2 Hypothèses fréquentiste et émotionnelle

Une première hypothèse qui pourrait expliquer cet avantage des visages joyeux dans les tâches de reconnaissance, bien que sans doute la moins étudiée, concerne la fréquence à laquelle nous les rencontrons dans la vie quotidienne. En ce sens, une étude de Calvo et Gutiérrez-García (2014) a montré qu'il existe une corrélation entre les scores de reconnaissance des expressions et la fréquence à laquelle nous sommes exposés à ces expressions. Ainsi, les visages joyeux sont les plus fréquents et les visages apeurés sont les moins fréquents. Les auteurs supposent que les expressions que nous observons plus souvent pourraient nous amener à construire un modèle visuel plus précis de leurs caractéristiques faciales et de leur structure, ce qui peut ensuite faciliter leur reconnaissance (voir Somerville et Whalen, 2006 pour des résultats similaires). D'autres hypothèses qui pourraient expliquer cet avantage en faveur des visages joyeux sont liées à la valence émotionnelle des expressions faciales. Par exemple, l'expression joyeuse est considérée comme l'expression la plus plaisante de toutes les expressions testées (Eisenbarth et al., 2008). Ainsi, cet avantage pourrait traduire un biais de traitement en faveur des stimuli positifs en général. Cette hypothèse est corroborée par quelques études dans lesquelles les différences physiques entre les expressions sont contrôlées par l'usage de visages

schématiques, qui montrent que les visages joyeux (des schémas avec une courbe en U) sont mieux identifiés que les visages tristes (des schémas avec une courbe en U inversée ; Leppänen et Hietanen, 2004 ; Song et al., 2017). Néanmoins, cette hypothèse à elle seule ne pourrait pas expliquer les différences entre les autres expressions. Par exemple, les visages apeurés peuvent être moins bien reconnus sans être considérés comme les moins plaisants (Eisenbarth et al., 2008). Toujours en lien avec la valence émotionnelle, l'avantage de la joie pourrait aussi venir du fait qu'elle est souvent la seule expression positive, comparée à plusieurs expressions négatives. Néanmoins, cette hypothèse n'expliquerait toujours pas les différences entre les autres expressions. De plus, l'avantage des visages joyeux se retrouve dans des études dans lesquelles seulement une expression négative est utilisée (voir par exemple Calvo et Nummenmaa, 2009 ; Leppänen et Hietanen, 2004).

1.5.2.3 Hypothèse physique

Sans nécessairement écarter toute implication de la valence émotionnelle ou de la fréquence d'apparition, les caractéristiques physiques des visages joyeux pourraient aussi être à l'origine de leur meilleure reconnaissance. Par exemple, une étude de Calvo et Nummenmaa (2011) a mis en évidence une contribution plus forte de la saillance de la bouche que de la valence émotionnelle dans la vitesse de discrimination des expressions faciales. En fait, entre plus d'être distinctif, le sourire des visages joyeux est saillant, ce qui pourrait faciliter l'accès à l'information utile au décodage de l'expression joyeuse (Calvo et al., 2012 ; Calvo et Nummenmaa, 2008). Dans plusieurs études, la région de la bouche a été identifiée comme particulièrement importante pour la catégorisation des visages joyeux (Beaudry et al., 2014 ; Bombari et al., 2013 ; M. L. Smith et al., 2005). Lorsqu'elle est cachée, les performances diminuent considérablement, tandis que, lorsque ce sont les yeux qui sont cachés, les performances ne changent pas (Beaudry et al., 2014). Dans la revue de la littérature de Calvo et Numennmaa (2016), les auteurs proposent que des caractéristiques morphologiques saillantes et distinctives facilitent la reconnaissance des expressions. Le sourire des visages joyeux capterait rapidement l'attention, et pourrait être utilisé comme un raccourci pour identifier l'expression la joie. Aussi, grâce à ses caractéristiques physiques, l'expression joyeuse pourrait être basée sur un traitement plus global que les autres expressions, dont la reconnaissance nécessiterait une analyse plus locale (Srinivasan et Hanif, 2010). Cette hypothèse d'une origine physique de l'avantage des visages joyeux dans les tâches de reconnaissance est aussi mise en avant par des travaux computationnels, qui utilisent des réseaux de neurones artificiels.

1.5.3 Les réseaux de neurones artificiels, des outils pour dissocier les processus perceptifs et émotionnels ?

Les réseaux de neurones artificiels sont des modèles de calcul qui s'inspirent de la structure et de la dynamique de réseaux de neurones biologiques. Ils comptent parmi les modèles les plus puissants pour la reconnaissance d'objets ou d'émotions sur des images. Ils sont basés sur plusieurs unités connectées, chaque unité modélisant un neurone, et chaque

connexion modélisant une synapse. Durant une phase d'apprentissage, les réseaux vont pouvoir apprendre à différencier des images selon leur catégorie (par exemple des images contenant des visages joyeux et des images contenant des visages apeurés) en modifiant le poids des connexions entre leurs neurones. Ensuite, ils seront capables de catégoriser de nouvelles images (par exemple de nouvelles images de visages) en généralisant les connaissances acquises lors de la phase d'apprentissage. La plupart du temps, ces réseaux de neurones sont utilisés dans le but d'améliorer les performances de classification dans des domaines appliqués, mais ils peuvent aussi être utilisés dans la recherche, pour mieux comprendre le comportement humain. Ainsi, certaines études se sont intéressées à la reconnaissance des émotions chez les réseaux de neurones artificiels, dans le but de mieux comprendre la reconnaissance des émotions chez les humains (Dailey et al., 2002 ; Li et Cottrell, 2012 ; Mermillod et al., 2010 ; Mermillod et al., 2019 ; Mermillod et al., 2009). L'intérêt d'utiliser des réseaux de neurones artificiels dans l'étude des expressions faciales est que, dans ces modèles, les expressions sont traitées sur la base des propriétés physiques uniquement. Le traitement émotionnel n'est donc pas pris en compte, et il est possible d'estimer si un traitement comparable des expressions est effectué par des humains et par des machines qui n'ont pas accès au contenu émotionnel des images. Une similarité entre la reconnaissance d'émotions chez les humains et les réseaux de neurones a été mise en évidence par certaines études (Dailey et al., 2002 ; Li et Cottrell, 2012). Ces études ont utilisé un modèle, appelé EMPATH, composé de trois étapes. D'abord, les images sont filtrées par un banc de filtres de Gabor, afin de modéliser les réponses des cellules de V1 (c'est-à-dire des réponses sensibles à une bande de fréquences et une orientation particulière). Ensuite, une méthode de réduction des données est appliquée et les données sont catégorisées selon six catégories, correspondant aux six émotions de base, par un réseau de neurones très simple. Les résultats ont montré que, aussi bien pour les humains que pour le réseau, la peur était l'émotion la plus difficile à reconnaître, tandis que la joie était l'émotion la plus facile à reconnaître. Globalement, ces résultats suggèrent que la reconnaissance des émotions peut reposer uniquement sur les propriétés physiques des images et par conséquent sur les processus perceptifs qui les intègrent. Aussi, en dehors d'une comparaison directe des performances, Mermillod et al. (2010) ont montré, en utilisant un modèle similaire, que les BFS permettraient de meilleures performances pour discriminer des émotions que les HFS. Les BFS constitueraient donc le signal utile à la reconnaissance d'expressions émotionnelles.

1.5.4 Attributs diagnostiques

Ainsi, certaines expressions sont plus faciles à reconnaître que d'autres, et les caractéristiques physiques associées à chaque expression pourraient expliquer ces différences de performances. Cela nous amène naturellement à nous demander quelles sont les régions du visage les plus diagnostiques, c'est-à-dire les plus utiles, dans la reconnaissance des expressions faciales. Bien souvent, l'attribution de l'attention aux différentes zones du visage est étudiée par l'intermédiaire des mouvements oculaires. Plusieurs études ont ainsi

mis en évidence l'importance de la région des yeux dans le traitement des visages en général. Par exemple, la zone des yeux est la zone la plus fixée lorsque l'on doit se forger une impression sur un visage (Janik et al., 1978) ou explorer librement des images de visages (Cangöz et al., 2013 ; Hernandez et al., 2009 ; Scheller et al., 2012). De plus, cette observation semble assez indépendante de la tâche à accomplir. Par exemple, Scheller et al. (2012) ont étudié les mouvements oculaires de participants dans des tâches de catégorisation d'émotions, de catégorisation de genre et d'exploration libre. Ils ont observé que, dans chacune de ces tâches, les participants fixaient la région des yeux beaucoup plus longtemps que les autres régions du visage. Il y avait néanmoins des différences en fonction de l'expression du visage. La région des yeux était fixée d'autant plus souvent pour les visages apeurés ou neutres, et la région de la bouche était plus souvent fixée pour les visages joyeux. Ces observations se retrouvaient indépendamment de la zone de présentation des images (en haut, en bas ou au milieu de l'écran), ce qui a conduit les auteurs à suggérer que les caractéristiques physiques des expressions modulent l'attention indépendamment de la tâche et de l'emplacement du visage. Des résultats similaires ont été observés dans une autre étude, dans laquelle les participants devaient juger la valence de visages émotionnels. Les yeux étaient fixés plus longtemps que les autres régions, mais pour les visages joyeux la différence de temps de fixation entre les yeux et la bouche était amoindrie (Eisenbarth et Alpers, 2011). Aussi, dans une tâche de discrimination d'expressions faciales, Schurgin et al. (2014) ont mis en évidence une focalisation plus importante (c'est-à-dire des temps de fixation plus élevés) sur les lèvres pour les visages joyeux, et sur les yeux pour les visages tristes, en colère, honteux ou apeurés.

En utilisant l'oculométrie, ces premières études suggèrent que les yeux sont particulièrement importants pour décoder les expressions faciales, et que la zone de la bouche est particulièrement importante pour décoder une expression joyeuse. Afin d'identifier les régions utiles à la reconnaissance des émotions de base, d'autres études ont utilisé des techniques qui consistent à masquer certaines parties du visage pour voir à quel point leur absence impacte les performances de catégorisation. Notamment, M. L. Smith et al. (2005) ont utilisé une méthode appelée *Bubbles* pour identifier les pixels les plus pertinents lors de la catégorisation d'émotions de base. Plus précisément, les participants devaient identifier les émotions présentées sur un visage, tandis que certaines parties du visage étaient masquées par des bulles. Les résultats obtenus sont représentés sur la Figure 1.10. Sur cette figure, on peut voir quels sont les pixels qui ont été associés à de meilleures performances de catégorisation. Par exemple, l'information située aux alentours de la bouche est particulièrement utile à l'identification de l'expression de la joie, tandis que l'information située au niveau des yeux est particulièrement utile à l'identification de l'expression de la peur. Similairement, dans le cadre de la catégorisation de visages neutres et expressifs (joyeux uniquement), la méthode des *Bubbles* a mis en évidence une importance particulière de la région de la bouche (Gosselin et Schyns, 2001 ; Schyns et al., 2002). Dans une autre étude, Wegrzyn et al. (2017) ont aussi présenté des visages avec différentes parties masquées à des participants, dans le but d'identifier les régions les plus utiles à l'identification de chaque émotion. Leurs résultats sont présentés sur la Figure

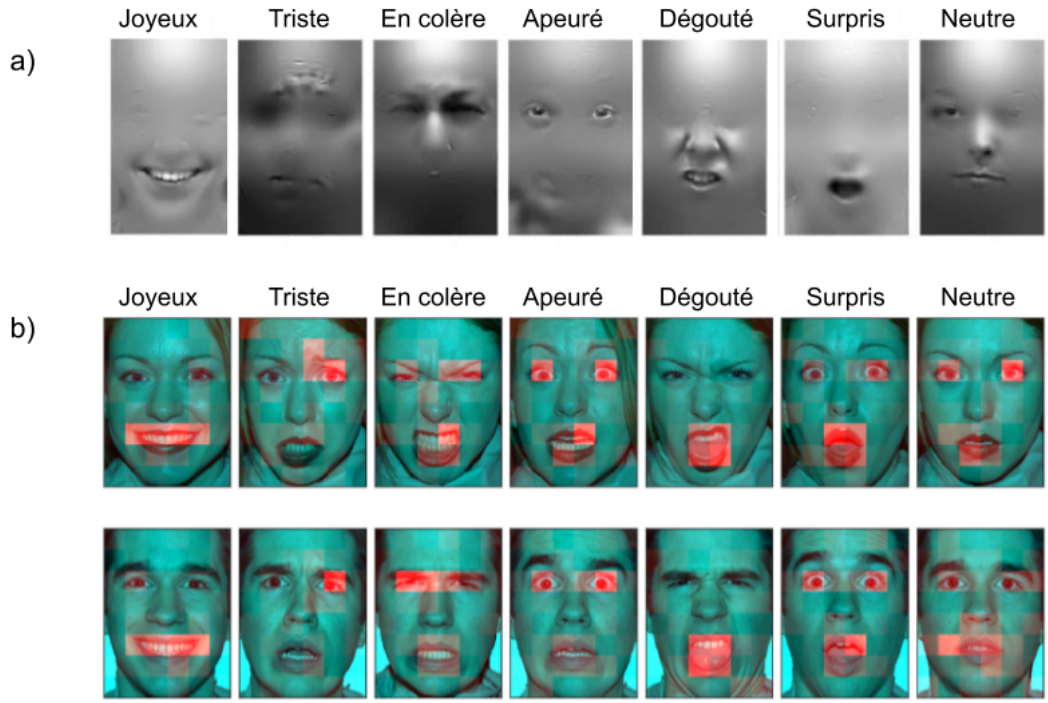

Figure 1.10 – Exemple de résultats illustrant les régions utiles pour catégoriser les émotions de base. **(a)** Extrait de M. L. Smith et al. (2005). **(b)** Extrait de Wegrzyn et al. (2017). Les tuiles sont colorées en fonction de leur importance, une forte importance étant associée à une tuile rouge et une faible importance à une tuile verte.

1.10, où les régions les plus importantes sont colorées en rouge. On retrouve encore une fois une importance particulière de la bouche dans le décodage d'une expression joyeuse. Pour la peur, les yeux sont les plus importants, mais la bouche est aussi utile, dans une moindre mesure. Ainsi, les observateurs semblent déployer différentes stratégies pour reconnaître les émotions de base en fonction de leurs attributs diagnostiques.

1.5.5 Usage flexible des fréquences spatiales

Tandis que la perception précoce des visages semble reposer sur un traitement des BFS, plusieurs études ont mis en avant une utilisation flexible des fréquences spatiales dans le cadre du décodage des expressions faciales. En effet, différentes bandes de fréquences seraient utilisées selon l'expression faciale à décoder (Kumar et Srinivasan, 2011 ; Morrison et Schyns, 2001 ; Schyns et al., 2009 ; F. W. Smith et Schyns, 2009) ou la tâche à effectuer (Schyns et Oliva, 1999 ; M. L. Smith et Merlusca, 2014). Par exemple, Schyns et Oliva (1999) ont constaté que, lorsqu'il était demandé à des participants de catégoriser des expressions faciales comme étant en colère, joyeuses ou neutres, les BFS étaient plus souvent utilisées. En revanche, lorsque la tâche était d'indiquer si le visage était expressif ou neutre, les HFS étaient plus souvent utilisées. Ces résultats ont été obtenus en utilisant des images hybrides, constituées des BFS d'un visage (par exemple un visage apeuré) et des HFS d'un autre visage (par exemple un visage joyeux), et illustrent bien l'importance de la tâche dans l'utilisation des fréquences spatiales pour décoder les expressions. Concernant

la différence entre les expressions, elle peut émerger du fait que les parties diagnostiques au décodage de chaque expression sont transmises par différentes bandes des fréquences. Ainsi, en utilisant la méthode des *Bubbles* avec des images filtrées, F. W. Smith et Schyns (2009) ont pu mettre en évidence les zones utiles pour la catégorisation des émotions de base dans différentes bandes de fréquences. Les yeux grands ouverts, diagnostiques aux visages apeurés, étaient plutôt transmis par les HFS, tandis que les régions utiles pour la joie étaient discernables à la fois en HFS et en BFS. Toujours afin d'identifier les fréquences utiles au décodage de différentes expressions, Kumar et al. (2011) ont demandé à des participants d'identifier des expressions faciales joyeuses ou tristes, présentées en BFS, en HFS ou non filtrées. Ils ont observé que la catégorisation de l'expression de la joie était plus efficace en BFS, tandis que la catégorisation de l'expression de la tristesse était plus efficace en HFS.

Afin de mieux comprendre comment les fréquences spatiales sont utilisées pour discriminer les expressions, Schyns et al. (2009) ont combiné la méthode des *Bubbles* à un enregistrement de l'activité en électroencéphalographie (EEG). Ils proposent qu'un premier traitement de l'information soit focalisé sur la région des yeux (à partir de 140 ms). Ensuite, un second traitement, plus global, serait mis en place (à partir de 156 ms). Pour finir, un traitement local des visages aurait lieu (à partir de 180 ms) et se concentrerait sur les fréquences utiles au décodage de l'expression présentée. Par conséquent, l'utilisation des fréquences spatiales ne serait pas un processus fixe, mais dépendrait de l'échelle des régions diagnostiques, en considérant à la fois les contraintes de la tâche et la configuration du visage.

1.5 L'expression faciale des émotions : reconnaissance et attributs - Points clés

- De manière générale, les tâches de catégorisation ont mis en évidence une bonne reconnaissance de la joie et une moins bonne reconnaissance de la peur. Cela pourrait s'expliquer par les propriétés physiques qui distinguent les visages joyeux des autres visages.
- Les réseaux de neurones artificiels peuvent être utilisés comme outils pour différencier les processus émotionnels des processus perceptifs.
- La reconnaissance des expressions faciales se base majoritairement sur l'information des yeux et de la bouche. La bouche est particulièrement utile pour reconnaître la joie, tandis que pour la peur les yeux ont plus d'importance.
- Les fréquences spatiales seraient utilisées de manière flexible pour décoder les expressions, en fonction de la tâche ou de l'expression.

1.6 Une capture "automatique" de l'attention par les visages émotionnels ?

Il est plutôt bien établi qu'un stimulus visuel avec un contenu émotionnel, en comparaison avec un stimulus visuel dont le contenu est neutre, va bénéficier d'un

traitement privilégié au sein du système visuel. Ainsi, il aura tendance à attirer plus facilement l'attention et les mouvements oculaires, et son analyse sensorielle sera améliorée (Vuilleumier, 2015). En ce sens, plusieurs études ont montré que, lorsqu'une scène visuelle émotionnelle et une scène visuelle neutre sont présentées simultanément de chaque côté d'un écran, la probabilité que l'image émotionnelle soit fixée en premier est plus élevée (D'Hondt et al., 2016 ; Koller et al., 2019 ; Nummenmaa et al., 2006). Dans la section précédente, nous nous sommes intéressés à la reconnaissance des expressions faciales. Dans cette section, nous nous intéresserons plus particulièrement aux études qui mettent en avant une capture de l'attention ou du regard par des expressions faciales émotionnelles (en comparaison avec des expressions faciales neutres). En effet, nombreuses sont les études comportementales qui, à travers différents paradigmes expérimentaux, ont mis en avant une priorité des visages émotionnels dans l'attribution des ressources attentionnelles (pour des revues, voir par exemple Carretié, 2014 ; Mulckhuyse, 2018 ; Palermo et Rhodes, 2007 ; Yiend, 2010). Nous n'en présenterons ici qu'une partie, en nous concentrant sur certains paradigmes pertinents dans le cadre de ce travail de thèse. Nous nous questionnerons alors sur l'aspect automatique [5] de cette capture.

1.6.1 Évidences issues des paradigmes de recherche visuelle

Le paradigme de recherche visuelle, dans le cadre du traitement des expressions faciales, consiste à présenter un ensemble de visages à des participants, et à leur demander de trouver le plus rapidement possible celui dont l'expression est discordante. Cette méthode a été utilisée dans de nombreuses études, et peut s'apparenter aux situations de la vie quotidienne dans lesquelles nous cherchons un objet pertinent parmi d'autres objets moins pertinents. Bien que, la plupart du temps, les participants répondent manuellement (les mouvements oculaires ne sont pas enregistrés), cette tâche permet de mettre en évidence les expressions faciales qui attirent plus facilement l'attention, que ce soit lorsqu'elles sont la cible de la recherche ou lorsqu'elles sont des distracteurs. Les études utilisant ce type de paradigme ont mis en évidence de manière consistante un avantage des visages émotionnels par rapport aux visages neutres, qui se traduit par des temps de réaction plus faibles lorsqu'il faut trouver un visage émotionnel parmi des visages neutres que l'inverse (pour une revue de la littérature sur la recherche visuelle d'expressions faciales, voir Frischen et al., 2008). Néanmoins, lorsqu'il s'agit d'identifier précisément quelle expression émotionnelle attire plus facilement l'attention, les résultats sont plus hétérogènes. En effet, il y a environ autant d'études qui rapportent un avantage des visages en colère (par exemple Coelho et al., 2010 ; E. Fox et al., 2000 ; Öhman et al., 2001 ; Schubö et al., 2006), que d'études qui rapportent un avantage des visages joyeux (par exemple D. V. Becker et al., 2011 ; Calvo et Marrero, 2009 ; Horstmann et al., 2012 ; Horstmann et Becker, 2020).

Ainsi, plusieurs facteurs ont été identifiés pour expliquer les différences obtenues entre les études (Frischen et al., 2008 ; Lundqvist et al., 2015). D'abord, les résultats

5. La notion d'automaticité fait référence ici à un processus rapide, inconscient, inévitable et peu coûteux en termes de ressources cognitives (Palermo et Rhodes, 2007 ; Yiend, 2010).

semblent très dépendants du type de stimuli utilisés (Juth et al., 2005 ; Savage et al., 2013). Par exemple, Juth et al. (2005) ont observé un avantage des visages joyeux avec des photographies et un avantage des visages en colère avec des visages schématiques. Aussi, Savage et al. (2013) ont observé des différences entre les photographies en fonction de la base d'images utilisée. La visibilité des dents est aussi un facteur important. Ainsi, lorsque des visages joyeux avec une bouche ouverte sont comparés à des visages en colère avec une bouche fermée, un avantage des visages joyeux est observé, alors que c'est l'inverse qui est observé lorsque la bouche est ouverte pour les visages en colère et fermée pour les visages joyeux (Horstmann et al., 2012). Le genre de la cible peut aussi influencer les résultats. Ainsi, Öhman et al. (2010) ont observé un avantage de la joie lorsque la cible était un visage féminin, et un avantage de la colère lorsqu'elle était un visage masculin (les résultats étaient aussi dépendants du nombre d'individus différents présentés dans l'expérience, et l'avantage de la colère pour les visages masculins était seulement observé lorsque le même individu était présenté dans chaque essai). Aussi, en reprenant les résultats de plusieurs études et en les comparant aux scores d'*arousal*[6] associés aux images utilisées, Lundqvist et al. (2014) ont mis en évidence une corrélation positive entre ces variables. Au final, il semble que les performances peuvent être expliquées à la fois par des facteurs émotionnels et des facteurs physiques (Lundqvist et al., 2015).

La plupart des études citées précédemment utilisent des réponses manuelles. Néanmoins, certaines études ont enregistré les mouvements oculaires des participants dans des tâches de recherche visuelle (S. I. Becker et al., 2017 ; Calvo et Nummenmaa, 2008 ; Devue et Grimshaw, 2017 ; Horstmann et Becker, 2020 ; Reynolds et al., 2009). Par exemple, en 2008 Calvo et Nummenmaa ont évalué la détection de visages émotionnels joyeux, surpris, dégoûtés, apeurés, en colère, ou tristes parmi des visages neutres en utilisant une réponse manuelle, mais aussi un enregistrement des mouvements oculaires. Ils ont observé que les visages joyeux, mais aussi dans une moindre mesure, surpris et dégoûtés, étaient localisés et fixés plus tôt. La saillance visuelle offrait une meilleure prédiction des performances que la valence émotionnelle.

1.6.2 Évidences issues des paradigmes de choix saccadique

À notre connaissance, deux groupes de recherche ont utilisé le paradigme de choix saccadique dans le cadre de la perception des expressions faciales. Premièrement, Bannerman et al. se sont intéressés à l'orientation de l'attention vers des visages émotionnels, en particulier apeurés, par rapport à des visages neutres. Dans une première étude, ils ont opposé deux visages, l'un émotionnel et l'autre neutre (Bannerman, Milders et Sahraie, 2009). Les participants devaient orienter leur regard vers le visage émotionnel ou neutre (réponse saccadique) ou indiquer la position du visage émotionnel ou neutre manuellement, en utilisant les boutons du clavier (réponse manuelle). Ils ont observé des temps de réponse plus faibles et des taux de réponses correctes plus élevés, lorsque

6. L'arousal d'un stimulus peut se définir comme un niveau d'excitation (Duffy, 1962), souvent associé l'intensité émotionnelle.

la cible était le visage émotionnel, mais seulement quand la réponse était saccadique. Ce pattern se retrouve aussi bien avec des visages schématiques (Expérience 1, visages émotionnels joyeux ou en colère) que des photographies (Expérience 2, visages émotionnels joyeux ou apeurés). Dans une troisième expérience, ils ont opposé des photographies de visages neutres, joyeux, ou apeurés à des contours vides de visages. Ils ont montré que les participants orientaient plus efficacement leur regard vers des visages apeurés que neutres. Cette étude souligne la sensibilité des mesures saccadiques aux effets émotionnels (en comparaison à une réponse manuelle) et témoigne d'une orientation de l'attention plus efficace vers des visages émotionnels, en particulier apeurés. Cet avantage des stimuli émotionnels a été reproduit avec des visages neutres et apeurés par la même équipe dans une autre étude, et a été étendu aux images dans lesquelles seulement le corps est présenté (position défensive ou neutre ; Bannerman, Milders, de Gelder et al., 2009).

À peu près au même moment, une deuxième série d'études a été menée par Calvo et al., et s'est particulièrement intéressée au décours temporel de la reconnaissance des visages joyeux. Dans une première étude, ils ont montré que, dans une tâche de choix saccadique, les participants étaient plus rapides pour détecter un visage joyeux par rapport à un visage triste, en colère, apeuré, surpris ou dégoûté (Calvo et Nummenmaa, 2009). Dans leur expérience, deux visages de différentes expressions (l'expression cible et un visage neutre) étaient simultanément présentés et les participants devaient faire une saccade vers le visage cible, défini verbalement au début de chaque essai. Ils ont observé une détection plus rapide lorsque la cible était un visage avec une expression joyeuse plutôt que n'importe quelle autre expression. Dans une autre étude en choix saccadique, les auteurs se sont intéressés à la discrimination entre un visage joyeux et un visage représentant une autre expression. Ils ont mis en relation les performances dans la tâche avec des facteurs physiques (par exemple la luminance, la saillance de la bouche ou des yeux) ou sémantiques (par exemple la valence émotionnelle). Seule une contribution de la saillance de la bouche a été observée (Calvo et Nummenmaa, 2011). Finalement, ces études témoignent de l'efficacité de la détection des visages joyeux et de l'importance de la saillance de la bouche dans cet effet.

1.6.3 Limites : une capture conditionnée par la tâche ?

Les études en recherche visuelle citées précédemment ont montré une capacité des expressions faciales émotionnelles à attirer l'attention plus que les expressions faciales neutres. Néanmoins, ces études se limitent au cadre d'un traitement explicite des expressions, du fait que celles-ci étaient toujours pertinentes pour la tâche. Si cette capacité des expressions émotionnelles à capturer l'attention est automatique, elle devrait se produire même lorsque les expressions ne sont pas pertinentes pour la tâche. Dans une étude de Becker et al. (2017), les participants devaient effectuer une saccade vers un visage masculin ou féminin présenté parmi des visages du sexe opposé (voir Figure 1.11). On a donc une tâche de recherche visuelle d'un visage en fonction de son genre. Les auteurs ont observé que les visages cibles étaient fixés plus rapidement lorsqu'ils avaient une expression émotionnelle (joie ou colère) plutôt que neutre. De plus, si la cible était un

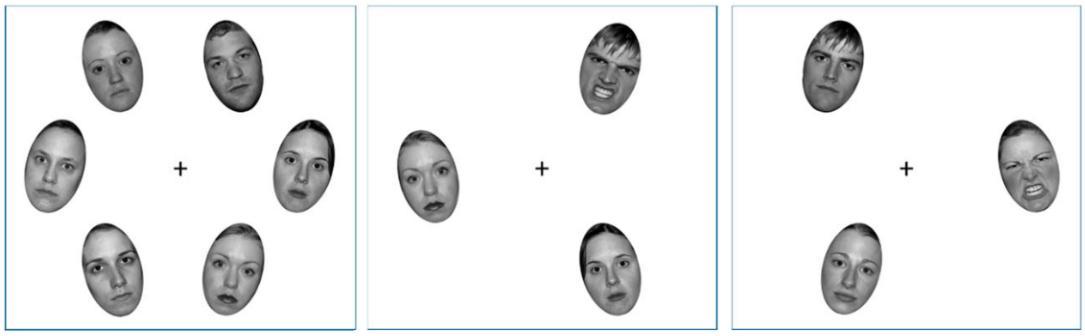

Figure 1.11 – Exemple de stimuli utilisés en recherche visuelle. Figure extraite de S. I. Becker et al. (2017).

visage neutre, elle était fixée moins rapidement en présence d'un distracteur émotionnel (en comparaison à une condition sans distracteur). Cet avantage des visages émotionnels était la plupart du temps similaire quelle que soit l'expression faciale (joie ou colère). Les auteurs ont néanmoins observé un avantage des visages en colère qui, en tant que distracteurs, attiraient la première fixation plus souvent que les visages joyeux (cet effet était seulement significatif avec 6 visages présentés et les auteurs suggèrent qu'il peut s'expliquer par le fait que les visages féminins en colère sont perçus comme plus masculins).

Bien que cette étude ait mis en évidence une capture de l'attention et du regard plus importante pour les visages émotionnels que neutres dans une tâche où les expressions ne sont pas pertinentes, d'autres études apportent une vision plus nuancée. Par exemple, dans une étude de Hunt et al. (2007) les participants devaient effectuer une saccade vers un visage schématique, en fonction de son expression (joyeuse, en colère ou neutre, Expérience 1) ou en fonction de son orientation (à l'endroit ou à l'envers, Expérience 2). Dans la moitié des essais, l'un des distracteurs était un visage avec une expression émotionnelle. Ils ont trouvé que la présence d'un distracteur, joyeux ou en colère, attirait l'attention (par l'intermédiaire d'une augmentation des temps de réaction), mais seulement lorsque les émotions étaient la cible de la recherche. Ainsi, les visages émotionnels n'attiraient pas l'attention dans la tâche de recherche de l'orientation. Ensuite, dans une autre étude, des participants devaient effectuer une saccade rapide vers un point de couleur parmi d'autres points dans une configuration circulaire (Devue et Grimshaw, 2017). Dans une configuration circulaire concentrique, à l'intérieur du cercle des points, des images d'objets non pertinents étaient présentées. L'une des images était un visage, neutre ou en colère, ou un papillon, tandis que les autres images étaient des photographies d'objets inanimés. Les résultats montraient que les deux types de visages capturaient la première saccade plus souvent qu'un papillon. Néanmoins, les visages neutres capturaient cette première saccade plus souvent que les visages en colère. Ils ont donc observé une capture de l'attention plus forte pour les visages que pour les autres objets, mais cet effet ne favorisait pas les visages émotionnels.

En résumé, les études en recherche visuelle avec des réponses oculaires, lorsque les expressions ne sont pas pertinentes pour la tâche, ne mettent pas systématiquement en

avant un traitement privilégié des visages émotionnels. Dans une revue de la littérature sur l'effet des émotions sur les mouvements oculaires, Mulckhuyse (2018) suggère que les saccades effectuées très rapidement (en moins de 200 ms) ne seraient pas sensibles au contenu émotionnel. Aussi, elle suggère que l'utilisation de stimuli de visages apeurés plutôt qu'en colère pourrait être plus appropriée pour mettre en avant une capture de l'attention par les visages émotionnels. En effet, les visages apeurés sont connus pour entraîner des réponses fortes et rapides au niveau de l'amygdale.

1.6 Une capture "automatique" de l'attention par les visages émotionnels ? - Points clés

- Les visages émotionnels sont détectés plus rapidement que les visages neutres.
- Certaines études témoignent d'un avantage des visages en colère, et d'autres d'un avantage des visages joyeux. L'hétérogénéité des résultats pourrait s'expliquer par le type de stimuli utilisé.
- L'aspect automatique de la capture de l'attention par les visages émotionnels est discutable. Les études en recherche visuelle dans lesquelles ils ne sont pas pertinents pour la tâche ne montrent pas systématiquement de différence entre les visages neutres et émotionnels.

1.7 Bases cérébrales du traitement des expressions faciales

Ainsi, certains stimuli semblent bénéficier d'un traitement privilégié au sein du système visuel. C'est le cas des visages, comparativement à d'autres objets, ou des visages émotionnels, comparativement à des visages neutres. Pour comprendre les effets que nous avons décrits dans les paragraphes précédents, il est important de comprendre la façon dont le cerveau traite l'information. Quelles sont les régions impliquées ? Quel est le rôle de chacune de ces régions et à quel moment interviennent-elles ? La neuroimagerie fonctionnelle, qui permet de caractériser l'activité cérébrale pour une tâche cognitive précise, est particulièrement indiquée pour répondre à ces questions. Dans cette section, nous commencerons par présenter les bases cérébrales du traitement des visages, qui se superposent partiellement avec celles du traitement des expressions faciales. Ensuite, nous reviendrons plus précisément sur les régions impliquées dans la perception des expressions faciales, ainsi que sur certains modèles influents de leur traitement. Bien que l'aspect automatique de la capture de l'attention par les visages émotionnels ne soit pas systématiquement mis en avant dans les études comportementales [7], nous verrons qu'il est au coeur de plusieurs modèles du traitement des émotions (pour des revues de la littérature, voir Diano et al., 2017 ; Mulckhuyse, 2018 ; Pessoa, 2010 ; Tamietto et de Gelder, 2010).

7. Ou, si ce traitement est automatique, il ne serait pas aussi rapide que certaines études le supposent. Il n'y a pas à notre connaissance de seuil précis pour caractériser si un processus est rapide ou non (Moors et De Houwer, 2006.). Plusieurs études considèrent des processus comme rapides/automatiques lorsqu'ils apparaissent aux alentours de 100 ms après l'apparition du stimulus (Palermo et Rhodes, 2007) ou interviennent de manière pré-attentive, avant la sélection attentionnelle consciente (Treisman, 1985)

1.7.1 Bases cérébrales du traitement des visages

1.7.1.1 Aires sélectives aux visages

Les études en IRMf ont permis de mettre en évidence les régions qui répondent spécifiquement aux visages, c'est-à-dire les régions qui sont plus fortement activées lorsque des visages sont présentés. Ainsi, trois régions sont classiquement associées au traitement des visages : l'aire fusiforme des visages (ou *fusiform face area* ; FFA), l'aire occipitale des visages (ou *occipital face area* ; OFA) et le sillon temporal supérieur postérieur (ou *posterior superior temporal sulcus* ; pSTS). Ces régions se trouvent aussi bien dans l'hémisphère droit que dans l'hémisphère gauche du cerveau, mais elles semblent s'activer avec plus d'intensité et plus souvent dans l'hémisphère droit (Haxby et Gobbini, 2011 ; Rossion et Lochy, 2021 ; Yovel, 2016). La Figure 1.12 propose une visualisation de ces différentes régions, qui furent les premières à avoir été identifiées et qui restent les plus étudiées. Néanmoins, plus récemment, d'autres zones sélectives aux visages ont été repérées dans des parties plus antérieures du cerveau. Plus précisément au niveau du lobe temporal antérieur, du sillon temporal antérieur supérieur et du gyrus frontal inférieur (Duchaine et Yovel, 2015). Plusieurs modèles ont été développés pour caractériser les différentes étapes du traitement des visages, ainsi que les aires cérébrales associées à chacune de ces étapes.

1.7.1.2 Modèle de Haxby et al. (2000)

L'un des modèles les plus influents du traitement des visages est celui proposé par Haxby et al., en 2000. Ce modèle est dérivé d'un autre modèle connu, celui de Buce et Young (1986), qui distinguait les processus impliqués dans la reconnaissance de l'identité de ceux impliqués dans la reconnaissance de l'expression faciale. D'après ce modèle, reconnaître une identité nécessite le décodage des aspects invariants d'un visage, c'est-à-dire de ce qui ne change pas en fonction des mouvements, des expressions, de l'éclairage, ou de l'angle de vue. Au contraire, reconnaître une expression faciale nécessite le décodage des variations dans le visage, qui ne sont pas pertinentes dans le cadre de la reconnaissance de l'identité. Ces variations englobent, par exemple les mouvements de la bouche ou l'écarquillement des yeux.

Dans ce cadre, Haxby et al. (2000) ont distingué un système central et un système étendu pour le traitement des visages. Le système central est constitué des aires sélectives aux visages : la FFA, l'OFA et le pSTS. Dans ce système, le traitement des aspects variants et invariants des visages est distingué. Ainsi, après une perception précoce des visages au niveau de l'OFA, la perception des mouvements des yeux et de la bouche se ferait au niveau du pSTS, tandis que la perception de l'identité se ferait au niveau de la FFA. Ce postulat a été établi à partir de résultats d'études en neuroimagerie, qui montrent que des changements d'identité, ou un focus de l'attention sur l'identité, sont associés à des activations neurales plus fortes au niveau de la FFA. Au contraire, des changements d'expressions, ou un focus de l'attention sur les mouvements des yeux, sont associés à des activations plus fortes au niveau du pSTS (Hoffman et Haxby, 2000 ; Puce et al., 1998). Au-delà du système central, le système étendu rassemble des aires cérébrales qui

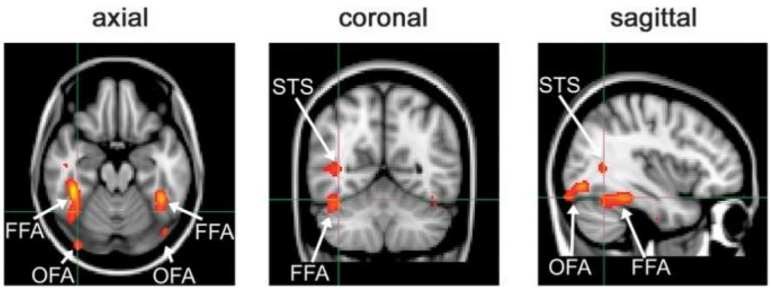

Figure 1.12 – Localisation des aires sélectives aux visages : la FFA, l'OFA et le pSTS. Figure extraite de Davies-Thompson et al. (2013).

participent à l'extraction d'informations supplémentaires, et implique la participation de systèmes neuraux qui ne sont pas nécessairement dédiés à la perception visuelle. Par exemple, le système étendu est composé de régions impliquées dans l'extraction des connaissances bibliographiques ou du contenu émotionnel, ainsi que de régions plus largement impliquées dans la compréhension de la parole, des intentions et des actions.

Bien que le modèle de Haxby et al. (2000) reste classique, différents auteurs en ont depuis proposé des révisions. Par exemple, Rossion (2008) suggère qu'il existe une voie directe entre le cortex visuel et la FFA, qui permettrait de détecter et catégoriser rapidement un visage sur la base d'un traitement holistique.

1.7.1.3 Modèle de Duchaine et Yovel (2015)

En 2015, Duchaine et Yovel proposent un modèle basé sur la distinction entre deux voies distinctes pour le traitement des visages : une voie dorsale et une voie ventrale. En comparaison avec le modèle de Haxby et al., de nouvelles aires sélectives aux visages sont intégrées, plusieurs voies d'accès au système des visages sont considérées (ainsi l'OFA ne serait plus le seul point d'entrée) et la fonction de certaines régions est redéfinie. Par exemple, la FFA serait aussi impliquée dans le traitement de certains aspects variants du visage, comme les expressions. La voie ventrale du traitement des visages comprend l'OFA, la FFA et l'aire des visages du lobe temporal antérieur, tandis que la voie dorsale comprend le pSTS, l'aire des visages du sillon temporal supérieur antérieur et celle du gyrus frontal inférieur. La voie ventrale serait majoritairement impliquée dans le traitement des aspects invariants du visage, comme l'identité, le sexe ou l'âge, mais contribuerait aussi à la reconnaissance des expressions faciales. L'OFA supporterait la perception précoce (dès 100 ms après l'apparition du visage) des parties du visage selon le point de vue. La FFA recevrait l'information de l'OFA mais aussi, comme le suggérait déjà Rossion, du cortex visuel. Elle supporterait le traitement holistique des visages. L'aire des visages du lobe temporal antérieur recevrait l'information de la FFA et l'OFA. Son rôle fonctionnel est encore peu connu, mais pourrait correspondre au traitement de l'identité. La voie ventrale serait impliquée dans l'analyse dynamique du visage, c'est-à-dire dans le traitement des aspects du visage qui changent rapidement, comme l'expression, le

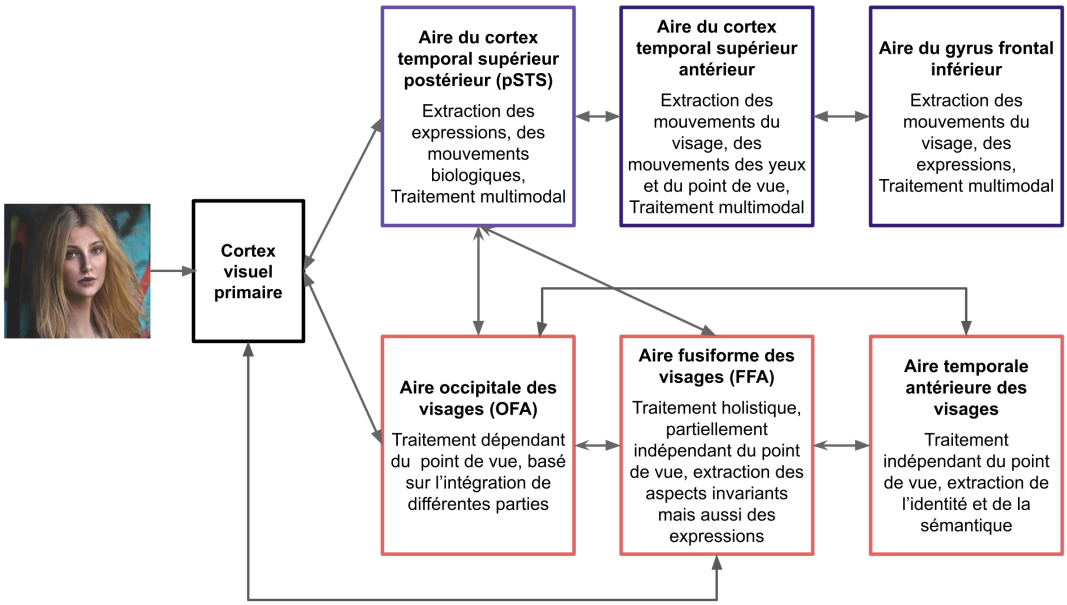

Figure 1.13 – Modèle de traitement des visages proposé par Duchaine et Yovel (2015). La voie ventrale du traitement du visage (en rouge) comprend l'OFA, la FFA et l'aire du visage du lobe temporal antérieur, tandis que la voie dorsale (en violet) comprend le pSTS, l'aire des visages du sillon temporal supérieur antérieur et celle du gyrus frontal inférieur. La voie ventrale est spécialisée dans l'extraction des caractéristiques de formes et la voie dorsale dans l'extraction de l'information dynamique.

regard et les mouvements de la bouche. Le pSTS recevrait les informations concernant les formes et les mouvements par l'intermédiaire du cortex visuel et les transmettrait à l'aire du visage du sillon temporal supérieur antérieur, ainsi qu'au gyrus frontal inférieur. Le rôle fonctionnel de ces dernières régions reste encore peu connu. Un résumé du modèle de Duchaine et Yovel est présenté sur la Figure 1.13.

1.7.2 Aires sélectives aux expressions faciales

1.7.2.1 Recouvrement avec les aires des visages, approche localiste et constructioniste

Les aires impliquées dans la discrimination des expressions du visage, c'est-à-dire les aires dont l'activité est accrue face à un visage émotionnel en comparaison à un visage neutre, sont très nombreuses (pour une méta-analyse des régions sélectives aux expressions faciales, voir Fusar-Poli et al., 2009 ; Liu et al., 2021). Parmi elles, nous retrouvons certaines aires sélectives aux visages. D'après le modèle de Haxby et al. (2000), c'est le pSTS qui serait impliqué dans la discrimination des expressions, tandis que la FFA serait impliquée dans l'identification. Néanmoins, de nombreuses études ont également observé des effets prononcés des émotions au sein de la FFA (C. J. Fox et al., 2009 ; Ganel et al., 2005 ; Kawasaki et al., 2012 ; Vuilleumier et Pourtois, 2007). C'est d'ailleurs ce qui a conduit Duchaine et Yovel (2015) à introduire cette précision dans leur modèle

du traitement des visages. Plus récemment, l'OFA a également été mise en avant pour sa capacité à distinguer les visages émotionnels des visages neutres (Liu et al., 2021). Ainsi, parmi les aires sélectives aux visages, l'OFA, la FFA et le pSTS ont chacune été associées à la perception des expressions faciales.

Dans le domaine des émotions, il y a longtemps eu un débat sur la question de savoir si elles sont représentées de manière localisée dans le cerveau (approche localiste), ou si elles partagent un réseau neural commun (approche constructioniste ; pour une revue de la question, voir Celeghin et al., 2017 ; Lindquist et al., 2012). L'approche localiste suppose que chacune des émotions de base est associée à un réseau qui lui est propre, distinct de celui des autres émotions. L'approche constructioniste suppose que les émotions émergent de l'interaction de vastes réseaux, impliqués dans des opérations assez générales. En utilisant des stimuli émotionnels variés, incluant des visages émotionnels, des méta-analyses ont mis en avant les régions les plus actives en fonction des émotions (Phan et al., 2002 ; Vytal et Hamann, 2010). Ainsi, la peur impliquerait particulièrement l'amygdale, le dégoût impliquerait particulièrement l'insula, la joie les ganglions de la base et la tristesse le gyrus cingulaire subcallosal. Dans une méta-analyse focalisée sur le traitement des expressions faciales émotionnelles, Fusar-Poli et al. (2009) ont analysé les différences d'activations entre les visages neutres et chacune des émotions de base (Figure 1.14). Ils ont mis en avant des activations plus prononcées pour des expressions de peur que des expressions neutres au niveau de l'amygdale (droite et gauche), du gyrus fusiforme et du gyrus frontal médian. Ils ont mis en avant des activations plus prononcées pour des expressions joyeuses que des expressions neutres toujours au niveau de l'amygdale (droite et gauche, mais les activations étaient moins prononcées que pour la peur) et du gyrus fusiforme, mais aussi au niveau du cortex cingulaire antérieur. Pour la colère, des activations plus prononcées sont observées au niveau de l'insula et du gyrus occipital inférieur. En résumé, ils ont mis en évidence des patterns différents, au moins partiellement séparables en fonction des expressions, mais aussi des structures communes à différentes expressions. Une méta-analyse plus récente a mis en avant plusieurs régions qui répondent plus fortement aux visages émotionnels que neutres de manière générale : l'amygdale (droite et gauche), le gyrus parahippocampique (droit et gauche), le gyrus occipital inférieur droit (incluant l'OFA), le gyrus occipital moyen gauche, la FFA gauche, le noyau ventral-latéral du thalamus gauche, et le gyrus frontal inférieur droit (Liu et al., 2021). Dans les paragraphes suivants, nous allons présenter plus en détail une région cruciale au traitement de l'information émotionnelle : l'amygdale. Nous nous concentrons ici sur cette région car son rôle est important dans le cadre des modèles qui seront présentés ensuite.

1.7.2.2 L'amygdale : une structure centrale

L'amygdale est une petite structure bilatérale en forme d'amande située dans le lobe temporal médian (notée Amy sur la Figure 1.14). Composée de plusieurs noyaux, elle est connectée à de nombreuses régions, corticales ou sous-corticales, comprenant notamment les aires sensorielles, le cortex orbitofrontal (ou *orbitofrontal cortex* ; OFC) et l'hippocampe

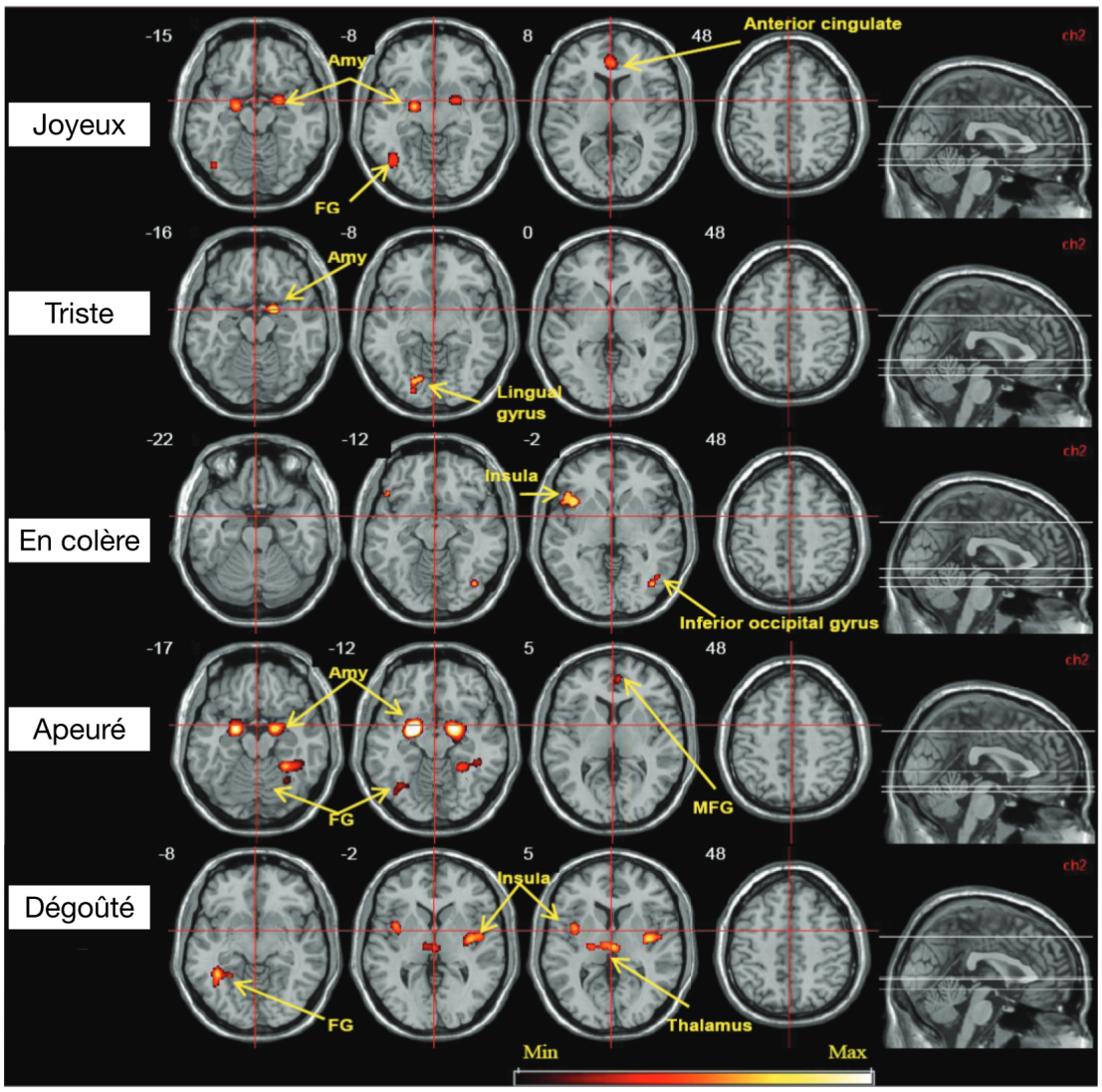

Figure 1.14 – Activations cérébrales en réponse à des visages émotionnels en comparaison à des visages neutres ($p < .001$). Les cartes sont issues de la méta-analyse de Fusar-Poli et al. (2009) qui rassemble les données de 105 études en IRMf. On observe des schémas au moins partiellement séparables, bien que les circuits impliqués ne soient pas totalement distincts.

(Janak et Tye, 2015 ; Pessoa, 2008, 2010 ; Robinson et al., 2010). Ces nombreuses connexions témoignent de son implication dans des tâches cognitives variées. À titre d'exemple, l'amygdale est impliquée non seulement dans l'évaluation de la pertinence émotionnelle (Adolphs, 2008 ; Davis et Whalen, 2001 ; Pessoa, 2010 ; Sander et al., 2003), mais aussi dans la formation de la mémoire (Phelps, 2004), l'attention (Holland et Gallagher, 1999), la prise de décision (Seymour et Dolan, 2008) et la régulation des réponses émotionnelles (Berboth et Morawetz, 2021 ; Frank et al., 2014 ; Goldin et al., 2008). Ainsi, l'amygdale semble impliquée dans la détermination de la nature d'un stimulus et la génération d'une réponse adaptée. De plus, elle n'agirait pas de manière isolée, mais servirait plutôt de nœud dans de multiples réseaux neuraux (Phelps, 2006 ; Šimić et al., 2021).

Initialement, les recherches sur l'amygdale ont cherché à lui attribuer une fonction

unique, et l'ont associée au traitement de la peur. En fait, bien que cette vision restrictive au traitement de la peur soit maintenant dépassée, de nombreuses études ont mis en évidence un lien fort entre le traitement de la peur et l'amygdale. Par exemple, une lésion de l'amygdale est associée à une déficience au niveau du conditionnement à la peur chez les rongeurs (LeDoux et al., 1990), de la reconnaissance d'expressions de peur (Adolphs et al., 1994 ; Calder, 1996) ou de la réponse émotionnelle face à des mots aversifs (Blanchard et Blanchard, 1972). Cette association entre l'amygdale et la peur est également appuyée par des études en neuroimagerie qui ont observé une activité plus forte de l'amygdale en réponse à des visages avec une expression faciale de peur, en comparaison à une expression de joie (Méndez-Bértolo et al., 2016 ; Morris et al., 1996 ; Mothes-Lasch et al., 2013), de dégoût (Phillips et al., 1998) ou neutre (Fusar-Poli et al., 2009 ; Vuilleumier et Sagiv, 2001). Néanmoins, sans remettre en question l'existence d'un lien entre l'amygdale et la peur, d'autres études ont souligné l'engagement de l'amygdale dans le traitement des visages neutres (Fusar-Poli et al., 2009), ou suggèrent qu'elle pourrait répondre de la même manière aux expressions négatives et positives en fonction de leur intensité émotionnelle (Bonnet et al., 2015 ; Fitzgerald et al., 2006 ; Garavan et al., 2001). Ainsi, plusieurs auteurs ont proposé que l'amygdale constitue un système de détection de la pertinence biologique plutôt qu'un module restreint au traitement de la peur (Adolphs, 2008 ; Pessoa, 2010 ; Sander et al., 2003 ; Weymar et Schwabe, 2016).

Finalement, certains auteurs font aussi un lien entre l'amygdale et la perception des différentes parties du visage. Par exemple, les lésions de l'amygdale sont associées à la fois à des déficits dans la reconnaissance de la peur, mais aussi à une exploration visuelle anormale des visages, qui se traduit par une baisse du nombre de fixations au niveau des yeux (Adolphs et al., 2005). Aussi, l'amygdale présente une activation plus forte en réponse à des yeux apeurés qu'à des yeux joyeux (Kanat et al., 2015 ; Whalen et al., 2004). Elle semble également particulièrement sensible à l'orientation du regard (Huijgen et al., 2015). Pour finir, l'amygdale pourrait s'activer très rapidement, et moduler l'activité de certaines aires visuelles via des connexions descendantes, afin d'optimiser l'encodage visuel des stimuli pertinents (Y. Chen et al., 2014 ; Furl et al., 2013 ; Vuilleumier et al., 2004). Par exemple, Vuilleumier et al. (2004) ont observé une activité plus forte en réponse à des visages apeurés que neutres dans les gyri occipital et fusiforme chez des sujets sains, mais pas chez des patients ayant subi une lésion de l'amygdale.

1.7.3 Modèles du traitement des expressions faciales

1.7.3.1 Modèle de Liu (2021) : un traitement cortical

Comme nous l'avons vu dans les paragraphes précédents, les modèles du traitement des visages intègrent déjà le traitement des expressions faciales. Par exemple, dans le modèle de Duchaine et Yovel (2015), les expressions faciales seraient distinguées au niveau de la FFA et du pSTS. Ces modèles décrivent un traitement des émotions et des visages que nous pouvons qualifier de cortical. L'information est transmise au cortex visuel par la voie rétino-géniculo-striée avant d'atteindre les régions spécifiques aux visages. À l'issue

d'une méta-analyse sur le traitement des visages émotionnels, Liu et al. (2021) ont proposé un modèle de traitement spécifique aux expressions faciales, qui se place dans la continuité des modèles du traitement des visages cités précédemment. Pour eux, le gyrus occipital joue un rôle important dans le traitement des expressions faciales. Au contraire, le rôle du pSTS dans le traitement des expressions faciales serait plus limité. Il concernerait le traitement de l'information dynamique d'une manière générale, comme proposé par Duchaine et Yovel (2015), mais pas le traitement des expressions spécifiquement. Dans ce modèle, les auteurs ont dissocié un traitement perceptif, impliquant en particulier la FFA et l'OFA, d'un traitement dépendant de la tâche, impliquant en particulier le gyrus frontal. Pour finir, les auteurs intègrent dans leur modèle des connexions entre les voies dorsale et ventrale du traitement des visages et des régions limbiques telles que l'amygdale et le gyrus parahippocampique. Pour eux, l'amygdale pourrait recevoir directement l'information issue des principales régions sélectives aux expressions : l'OFA, la FFA, le pSTS ainsi que les régions frontales. Une visualisation de leur modèle est présentée sur la Figure 1.15.

1.7.3.2 Modèles en double voie : un traitement cortical et sous-cortical

Le modèle de Liu et al. (2021) est à notre connaissance le modèle le plus récent du traitement des expressions faciales. Il caractérise un traitement cortical, faisant nécessairement intervenir la voie rétino-géniculo-striée. Néanmoins, en parallèle à ce traitement cortical, plusieurs auteurs suggèrent qu'il existerait une autre voie de traitement, que nous appellerons ici voie sous-corticale, qui ne ferait pas intervenir le cortex visuel primaire.

L'un des premiers modèles du traitement des émotions est celui de Ledoux (1998), qui est inspiré de travaux sur le conditionnement à la peur chez les rongeurs. Au cours de ses travaux, Ledoux a montré que la réponse émotionnelle associée à des stimuli acoustiques émotionnels était préservée après une ablation du cortex auditif (LeDoux et al., 1984), mais altérée après une lésion de la partie latérale de l'amygdale (LeDoux et al., 1990) [8]. Dans son modèle, Ledoux (1998) suggère qu'il existe deux voies de traitement des émotions : une voie haute et une voie basse. La voie haute transmettrait l'information à l'amygdale en passant par le cortex sensoriel, et bénéficierait d'une haute résolution spatiale. La voie basse relierait quant à elle directement le thalamus et l'amygdale pour permettre une perception rapide du contenu émotionnel. Cette voie rapide aurait une faible résolution spatiale, mais permettrait à l'organisme de réagir de manière réflexive en présence d'un danger.

En 2002, Adolphs propose un premier modèle spécifique au traitement des expressions faciales, qui se place dans la continuité du modèle de Haxby (2000) sur le traitement des visages. Il y intègre de manière un peu plus précise le traitement des expressions faciales, en détaillant les aires impliquées à différentes échelles temporelles. Selon ce modèle, similairement à ce qui a été proposé par Ledoux (1998), deux voies de traitement peuvent être distinguées. D'abord, une voie sous-corticale qui contourne le cortex visuel et dont on suppose qu'elle se limite à un traitement automatique et grossier. Ensuite, une

8. Les stimuli acoustiques émotionnels étaient des sons qui avaient été préalablement associés à des chocs électriques, et les réponses émotionnelles se caractérisaient par des changements au niveau du rythme cardiaque, de la pression artérielle ou du comportement d'immobilisation.

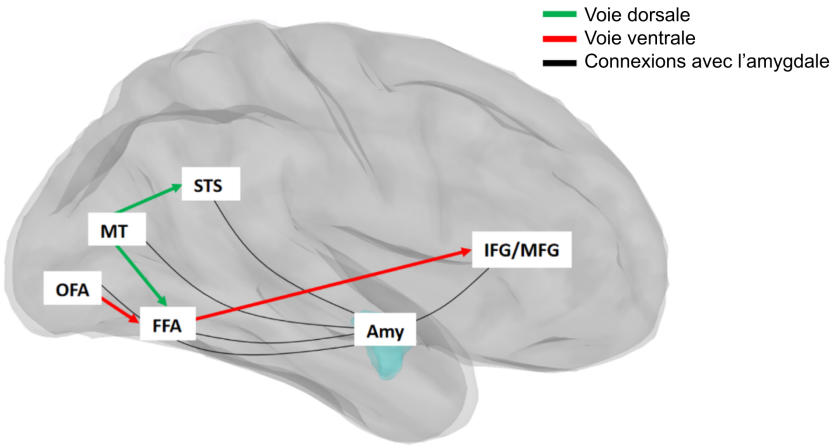

Figure 1.15 – Modèle de traitement des expressions faciales proposé par Liu et al. (2021). La voie ventrale est impliquée dans la perception des expressions faciales, tandis que la voie ventrale est impliquée dans la perception du mouvement.

voie corticale impliquant les aires visuelles occipitales et temporales, dont on suppose qu'elle permet l'émergence de représentations perceptives fines. Plus précisément, il y aurait après l'apparition d'un visage un premier traitement grossier qui permettrait l'encodage de la structure globale, et qui impliquerait la voie sous-corticale qui relie la rétine à l'amygdale, en passant par le CS et le pulvinar. L'amygdale contribuerait à une première évaluation de la pertinence émotionnelle du stimulus. Par la suite, environ 170 ms après l'apparition du stimulus, une représentation structurelle détaillée du visage est construite. La construction de cette représentation impliquerait notamment la FFA et le pSTS, dont les fonctions sont similaires à celles présentées dans le modèle de Haxby et al. (2000). L'amygdale et l'OFC interviendraient pour relier une représentation perceptive de l'expression à une connaissance conceptuelle de l'émotion. Cela passerait par une rétroaction vers le cortex visuel temporal et occipital, qui permettrait d'ajuster la catégorisation de l'expression faciale et d'allouer l'attention à certaines parties du visage plutôt que d'autres. L'amygdale et l'OFC vont aussi servir de lien avec diverses régions corticales, comme l'hippocampe pour déclencher les connaissances associées à l'expression faciale, ou des structures motrices pour générer une réponse émotionnelle.

Cette idée d'une séparation du traitement des expressions faciales en deux voies, l'une corticale et l'autre sous-corticale, est toujours d'actualité. Cependant, elle est au centre de nombreux débats, ce que nous aborderons dans la section suivante. Nous pouvons naturellement lier les modèles en double voie au traitement des fréquences spatiales. Contrairement à la voie corticale qui a accès aux HFS, la voie sous-corticale serait limitée au traitement des BFS. Dans une revue de la littérature, Tamietto et de Gelder (2010) reviennent sur les bases cérébrales de la perception non consciente des stimuli visuels émotionnels. Ils ont présenté un résumé assez complet des différentes voies visuelles et émotionnelles impliquées, présenté sur la Figure 1.16. Ainsi, il existerait deux voies

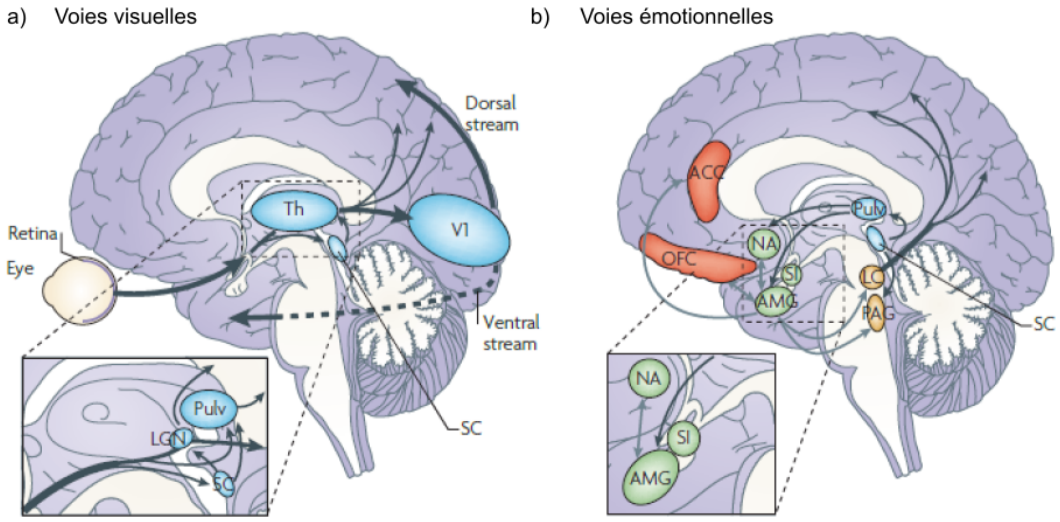

Figure 1.16 – Bases cérébrales de la perception des stimuli émotionnels. **(a)** Voies visuelles. La voie visuelle corticale, représentée par des flèches épaisses, part de la rétine et se projette vers V1 via un relais dans le CGL (LGN). À partir de V1, l'information visuelle atteint le cortex extrastrié le long de la voie ventrale et de la voie dorsale. Une minorité de fibres provenant de la rétine emprunte la voie sous-corticale, représentée par des flèches fines, pour atteindre le CS (SC) et le pulvinar (Pulv). Ceux-ci se projettent ensuite vers le cortex visuel extrastrié, en contournant V1. **(b)** Voies émotionnelles. Le système émotionnel comprend plusieurs zones, dont des structures sous-corticales comme l'amygdale (AMG), la substantia innominata (SI ; en vert), le noyau accumbens (NA) et les noyaux du tronc cérébral (en jaune). Il comprend aussi des zones corticales (représentées en rouge) comme l'OFC (OFC) et le cortex cingulaire antérieur (ACC). Les systèmes visuel et émotionnel sont largement interconnectés, en particulier au niveau sous-cortical avec les connexions entre l'amygdale et le CS par le pulvinar. Les flèches grises indiquent les connexions au sein du système émotionnel. Figure extraite de Tamietto et de Gelder (2010).

distinctes pour le traitement de l'information visuelle. D'abord, la voie visuelle corticale présentée au début de ce chapitre, qui part de la rétine et se projette vers le cortex visuel primaire via un relais dans le CGL. À partir du cortex visuel primaire, l'information visuelle atteint le cortex extrastrié le long de la voie ventrale et de la voie dorsale. Ensuite, la voie sous-corticale, empruntée par une minorité des fibres provenant de la rétine (environ 10%). Par cette voie, l'information va atteindre le CS et le pulvinar, puis elle va être transmise à l'amygdale et au cortex visuel extrastrié en contournant V1. Les systèmes visuel et émotionnel sont largement interconnectés, en particulier au niveau de l'amygdale qui est connectée à la fois au cortex visuel et au CS.

1.7.3.3 Le colliculus supérieur et du pulvinar : des relais vers l'amygdale ?

Ainsi, le CS et le pulvinar sont des régions importantes dans les modèles de traitement des expressions faciales, et des stimuli émotionnels en général, car ils constitueraient des points de relais pour atteindre l'amygdale. Le CS est connu pour sa fonction de guidage et de coordination dans les mécanismes d'orientation de l'attention, que ce soit par l'intermédiaire de la programmation de mouvements oculaires, comme nous l'avons

vu dans la section 1.3.3, de mouvements de la tête ou de réponses musculaires (White et Munoz, 2011). Les couches superficielles du CS reçoivent des projections directes des cellules magnocellulaires de la rétine et du cortex visuel primaire (Rodieck et Watanabe, 1993 ; Schiller et Malpeli, 1977 ; White et Munoz, 2011). Les neurones qui les composent encodent de manière précoce la saillance physique des objets de la scène, et forment ainsi une carte de saillance. Récemment, une étude chez les singes a observé que la saillance physique était représentée dans les couches superficielles du CS dès 65 ms après l'apparition du stimulus, avant qu'elle ne soit représentée dans les neurones de V1 (White, Kan et al., 2017). Les couches intermédiaires du CS intègrent des signaux provenant de multiples zones, corticales et sous-corticales (par exemple des zones frontales et pariétales), et se projettent vers des régions impliquées dans la préparation motrice (par exemple le tronc cérébral pour le déclenchement des saccades). L'activation des neurones des couches intermédiaires dépend des buts de l'observateur et forme une carte de priorité dont le pic d'activation correspond à la localisation de la cible d'une saccade (Bisley et Mirpour, 2019 ; White, Berg et al., 2017). En résumé, le CS, bien qu'il relie de nombreux circuits et desserve des fonctions à la fois sensorielles, motrices et cognitives, possède un rôle particulièrement important dans la détection de la saillance visuelle et la programmation de saccades.

Le pulvinar est le plus grand noyau du thalamus. L'une de ses fonctions est d'assister le traitement visuel en déplaçant l'attention vers les stimuli pertinents, et en éliminant les informations visuelles non pertinentes. Il reçoit notamment l'information de la rétine et des couches superficielles et intermédiaires du CS, et se projette vers de nombreuses aires visuelles, comme V1, V2, ou encore le sillon temporal supérieur (Soares et al., 2017). Il peut être divisé en plusieurs parties, et c'est la partie dorsale (composée du pulvinar médian et d'une partie du pulvinar latéral), qui reçoit l'information du CS et qui se projette vers l'amygdale, qui serait particulièrement impliquée dans le traitement rapide et automatique des stimuli émotionnels (Day-Brown et al., 2010 ; Jones et Burton, 1976 ; Soares et al., 2017).

1.7.3.4 Intérêt fonctionnel de la voie sous-corticale

En 2001, Öhman et Mineka proposent qu'il existe un module dans le cerveau pour le déclenchement et l'apprentissage de la peur. Selon les auteurs, ce module émergerait d'un circuit cérébral dédié, centré sur l'amygdale, qui n'est autre que le circuit sous-cortical qui a été présenté dans le paragraphe précédent. Dans ce contexte, la voie sous-corticale serait le fruit de l'évolution, sa fonction étant utile à la survie des espèces en général. Le module pour la peur supporté par la voie sous-corticale aurait plusieurs caractéristiques particulières. D'abord, il serait activé préférentiellement dans des contextes aversifs, par des stimuli qui sont pertinents dans une perspective évolutive et en lien avec la peur. Ensuite, son activation en réponse à ces stimuli serait automatique, et échapperait au contrôle cognitif. Ici, ce module n'est alors pas exclusivement associé aux expressions faciales. En effet, bien que les auteurs mettent en lien direct ce module et un avantage des visages en colère dans des tâches de recherche visuelle, il s'activerait en réponse à tous les

stimuli qui représentent une menace. Notamment, les animaux, surtout les prédateurs, ou les stimuli qui font l'objet de phobies. Cette vision d'un module pour la peur a été très influente. Depuis, on attribue plus généralement à la voie sous-corticale une fonction dédiée à la détection et la génération de réponses lorsque la survie exige une action rapide (Diano et al., 2017 ; Soares et al., 2017). Comme nous l'avons vu dans la section 1.1.3, il existerait au sein du système visuel une voie sous-corticale, impliquée dans vision non consciente. Cette hypothèse est notamment basée sur des études sur des patients souffrant de cécité corticale, pouvant toujours répondre à certains stimuli présentés dans leur champ visuel aveugle. Selon certaines études, les stimuli similaires aux visages sont particulièrement bien perçus par ces patients (Vuilleumier, 2000 ; Vuilleumier et Sagiv, 2001).

1.7.4 Les potentiels évoqués comme indices du décours temporel du traitement des expressions faciales

Comme nous l'avons vu dans la section 1.4.2, certaines tâches comportementales, comme la tâche de choix saccadique, permettent de rendre compte du décours temporel du traitement des visages. Un autre moyen d'obtenir des informations sur la dynamique temporelle du traitement des visages est l'EEG, qui permet d'enregistrer les courants électriques créés par l'activité neuronale postsynaptique par le biais d'électrodes posées sur le cuir chevelu. Classiquement, la composante N170 est associée à l'identification consciente d'un visage. C'est une onde négative, qui apparaît entre 120 et 200 ms après l'apparition d'un visage, et qui atteint son pic d'activité aux alentours de 170 ms. Elle ne serait pas simplement déclenchée par les caractéristiques physiques du stimulus, mais dépendrait plutôt de notre connaissance de ce qu'est un visage. Néanmoins, la détection des visages pourrait, au moins en partie, être basée sur des informations statistiques extraites plus tôt, au niveau de la composante P1, une onde positive qui atteint son pic aux alentours de 100 ms (Ganis et al., 2012 ; Rossion et Caharel, 2011 ; pour des revues de la littérature sur les potentiels évoqués à l'apparition d'un visage, voir Rossion, 2014 ; Yovel, 2016). Ainsi, l'IRMf et l'EEG fournissent des informations complémentaires sur l'activité cérébrale. Une étude de Sadeh et al. (2010) combine ces mesures, et a révélé une corrélation entre l'activité observée en EEG autour de 170 ms après l'apparition du visage et celle observée en IRMf au niveau de la FFA et du pSTS. L'activité observée en IRMf au niveau de l'OFA était corrélée avec celle observée en EEG de manière plus précoce, environ 110 ms après l'apparition du visage. Ce résultat souligne une précédence du traitement des visages opéré au niveau de l'OFA, sur celui opéré au niveau de la FFA et du pSTS.

Dans une récente revue de la littérature, Schindler et Bublatzky (2020) ont résumé les recherches sur les potentiels évoqués par la perception des expressions faciales. Selon cette revue, la composante P1 ne présente pas d'effet fiable des émotions. Il existe dans la littérature des études qui observent un effet des expressions faciales sur cette composante, mais autant d'autres qui n'en observent pas. En revanche, les émotions modulent l'activité de la composante N170, et de la composante EPN, une onde négative apparaissant entre 200 et 300 ms après l'apparition du stimulus dans les régions temporo-occipitales. La composante LPP, une onde négative qui apparaît entre 300 et 600 ms après l'apparition

du stimulus, dans les régions pariétales, est aussi impactée. L'EPN serait associée à un marquage perceptif précoce, et un reflet de la priorité attribuée à un stimulus en fonction de sa valence, qui pourrait se faire sur la base des informations perceptives. La LPP serait quant à elle le reflet de l'orientation de l'attention vers le stimulus, d'un étiquetage affectif et d'un encodage sémantique. Ces activités sont considérées comme des composantes typiques du traitement des émotions, et sont sensibles à divers stimuli visuels. À l'issue de leur revue, Schindler et Bublatzky (2020) suggèrent que le traitement des expressions faciales commence avec une identification visuelle précoce indépendante des ressources (P1), suivie d'un traitement configural précoce (N170) et d'une intégration des informations configurales pour l'évaluation de la pertinence (EPN). Si les ressources attentionnelles sont disponibles et allouées au contenu émotionnel pertinent, il en résulte une activité tardive soutenue (LPP).

1.7 Bases cérébrales du traitement des expressions faciales - Points clés

- Le réseau neural correspondant au traitement des visages est distribué. Il comprend notamment la FFA et l'OFA, qui permettraient de percevoir les expressions faciales.
- L'amygdale est une structure cruciale dans le traitement des expressions faciales. Particulièrement activée par des visages apeurés, elle serait généralement impliquée dans l'évaluation de la pertinence émotionnelle.
- Une voie sous-corticale reliant le CS, le pulvinar et l'amygdale pourrait être impliquée dans la détection rapide et non consciente des stimuli émotionnels, en particulier des visages apeurés.
- La voie sous-corticale traiterait l'information en BFS issue de certaines fibres magnocellulaires de la rétine. Elle opérerait en parallèle à la voie corticale, plus lente mais capable de traiter l'information en HFS.

1.8 Arguments en faveur de la voie sous-corticale et débats actuels

Nous avons vu dans les paragraphes précédents qu'une voie sous-corticale, reliant le CS à l'amygdale, permettrait de détecter rapidement les stimuli émotionnels, incluant les visages émotionnels. Bien que l'hypothèse de l'existence d'une telle voie ait été étudiée de manière extensive, elle fait l'objet de nombreux débats. Dans cette section, nous reviendrons dans un premier temps sur les données qui soutiennent cette hypothèse, puis nous aborderons les débats qui l'entourent. Nous terminerons cette section en présentant un bilan des différentes voies qui pourraient relier le système émotionnel et le système oculomoteur.

1.8.1 Des évidences pluridisciplinaires...

1.8.1.1 Évidences neuropsychologiques : le *blindsight* affectif

Les premières données en faveur de l'existence de la voie sous-corticale dans le cadre de la perception visuelle viennent de patients ayant subi des lésions du cortex visuel

primaire. En fait, en considérant que l'information visuelle est traitée uniquement par la voie corticale, nous nous attendrions à ce qu'une personne ayant subi une lésion du cortex visuel primaire gauche soit aveugle du côté droit de son champ visuel. Les études sur ce type de patients rapportent qu'ils déclarent en effet ne rien voir du côté aveugle de leur champ visuel (c'est-à-dire du côté opposé à leur lésion). Néanmoins, ces patients restent capables de répondre inconsciemment à des stimuli présentés dans cette zone aveugle (Pöppel et al., 1973 ; Sanders et al., 1974 ; Weiskrantz et al., 1974). Ce phénomène, connu sous le nom de vision aveugle ou *blindsight*, suggère qu'il existerait une ou plusieurs autres voies visuelles, qui contourneraient V1, et qui seraient impliquées dans la vision non consciente. Traditionnellement, ce rôle a été attribué à la voie sous-corticale décrite par les modèles en double voie. Cette hypothèse est corroborée par des données chez les singes, qui montrent qu'une lésion du CS après une lésion de V1 entraîne un déficit visuel complet (Pöppel et al., 1973 ; Sanders et al., 1974 ; Weiskrantz et al., 1974).

Il semble que le contenu émotionnel des stimuli, notamment celui transmis par les expressions faciales, soit particulièrement bien perçu par ces patients. On parle dans ce contexte de *blindsight* affectif (pour une revue de la littérature, voir Celeghin et al., 2015). Par exemple, la première étude à mettre en évidence ce phénomène a montré qu'un patient était capable de deviner si un visage qui lui était présenté dans son champ visuel aveugle exprimait de la joie ou de la peur (De Gelder et al., 1999). De plus, lorsque nous sommes face à une personne, nous avons tendance à faire correspondre nos propres expressions faciales à celles de cette personne, ce qui peut se mesurer par l'intermédiaire d'enregistrements de l'activité des muscles du visage. Et, des patients atteints de cécité corticale peuvent déclencher de telles réactions musculaires en réponse aux expressions faciales d'un visage, indépendamment du champ dans lequel il est présenté (Tamietto et al., 2009). Aussi, dans des tâches de recherche visuelle, ces patients présentent les mêmes biais que les participants sains dans le cadre de la recherche d'un visage émotionnel, quel que soit le champ de présentation de la cible (Lucas et Vuilleumier, 2008). Plus récemment, une étude a mis en évidence une réduction des temps de réaction dans une tâche qui consistait à discriminer l'orientation de patchs de Gabor présentés dans le champ intact, lorsque des visages apeurés étaient présentés simultanément dans le champ aveugle. Cet effet était néanmoins uniquement observé chez les patients présentant des lésions du cortex visuel gauche (Bertini et al., 2019).

Ainsi, ces études témoignent d'un traitement des expressions faciales qui se ferait indépendamment de la voie corticale. Une méta-analyse, menée par Celeghin et al., en 2019, a considéré l'ensemble des études en IRMf menées chez les patients atteints de cécité corticale. Les régions les plus actives comprenaient des structures sous-corticales, telles que le CS, le pulvinar et l'amygdale, ainsi que des zones extrastriées situées le long de la voie dorsale et de la voie ventrale. Les activations du CS, du pulvinar et de l'amygdale étaient particulièrement observées dans des études qui utilisaient des visages avec différentes expressions faciales. Ces résultats soutiennent ainsi l'hypothèse de l'implication de la voie sous-corticale dans le traitement non conscient des visages émotionnels.

1.8.1.2 Évidences neurophysiologiques : de l'IRM à l'EEG intracrânien

Les études sur les patients atteints de cécité corticale ne sont pas les seules à souligner l'implication de la voie sous-corticale dans la vision non consciente des stimuli émotionnels. Chez les sujets sains, une augmentation de l'activité de l'amygdale, du pulvinar et du CS a été mise en évidence, par exemple lorsque des visages apeurés étaient invisibles, c'est-à-dire masqués par un autre visage et non consciemment perçus, plutôt que visibles (Liddell et al., 2005 ; pour des résultats similaires avec des visages en colère, voir Morris et al., 1999). Aussi, plusieurs études ont mis en évidence l'implication du pulvinar dans le traitement des visages et des expressions faciales. Par exemple, une étude sur un patient ayant subi une lésion du pulvinar a mis en évidence chez ce patient une incapacité à reconnaître les expressions faciales apeurées (Ward et al., 2007). De plus, des enregistrements intracrâniens de l'activité cérébrale chez les singes ont montré que les neurones du pulvinar pouvaient répondre à différentes expressions de visages humains, avec des latences allant de 40 ms à plus de 300 ms (Maior et al., 2010). Ces latences hétérogènes sont cohérentes avec l'hypothèse d'une première réponse, rapide et grossière, qui précéderait une intégration plus tardive de l'information en provenance des régions corticales. Certaines études ont utilisé l'IRM de diffusion pour mettre en évidence les connexions anatomiques entre l'amygdale, le pulvinar et le CS (Koller et al., 2019 ; McFadyen et al., 2019 ; Rafal et al., 2015 ; Tamietto et al., 2012). En accord avec l'hypothèse sous-corticale, ces études ont montré que l'intensité des connexions entre ces régions (caractérisée par une mesure de la densité des fibres) était en lien avec l'intensité du biais d'orientation des mouvements oculaires vers des scènes négatives (Koller et al., 2019), ou avec les performances de reconnaissance des visages apeurés (McFadyen et al., 2019). De plus, ces connexions semblent accrues après une lésion du cortex visuel primaire (Tamietto et al., 2012).

Concernant le décours temporel de l'activation de ces régions, des enregistrements intracrâniens chez les singes ont mis en évidence une hausse précoce de l'activation des neurones du CS et du pulvinar en réponse à des stimuli ressemblant à des visages. Ces réponses apparaissaient dès 25 ms au sein du CS (Nguyen et al., 2014), et dès 50 ms au sein du pulvinar (Nguyen et al., 2013). Chez les humains, les enregistrements intracrâniens de l'activité des neurones de l'amygdale ont mis en évidence une activation plus forte pour les visages apeurés que neutres dès 140 ms (Pourtois et al., 2010), dès 75 ms (Méndez-Bértolo et al., 2016) ou dès 50 ms (W. Sato et al., 2011). Dans plusieurs études, l'encodage émotionnel opéré au niveau de l'amygdale précédait celui opéré au niveau du cortex visuel (Krolak-Salmon et al., 2004 ; Méndez-Bértolo et al., 2016 ; Pourtois et al., 2010 ; Sabatinelli et al., 2009). Une étude en EEG chez des patients ayant subi des dommages au niveau de l'amygdale a observé un impact de ces dommages sur l'activité aux alentours de 100-150 ms (composante P1), et plus tard aux alentours de 500-600 ms. Dans l'ensemble, ces données vont dans le sens d'une discrimination rapide des émotions au niveau de l'amygdale, supportée par la voie sous-corticale qui traverse le CS et le pulvinar.

1.8.1.3 Les basses fréquences spatiales : cruciales pour la détection précoce des visages émotionnels ?

Une particularité de ce réseau sous-cortical est qu'il serait basé uniquement sur une information grossière, en BFS. En effet, le CS reçoit l'information visuelle de la rétine depuis des cellules magnocellulaires (Schiller et Malpeli, 1977). Certaines études se sont intéressées à cette propriété de la voie sous-corticale, et ont montré une sélectivité du traitement émotionnel au contenu fréquentiel. Notamment, Vuilleumier et al. (2003) ont présenté des visages avec une expression neutre ou apeurée, sous différentes conditions de filtrage (non filtrés, en HFS ou en BFS) à des participants qui devaient discriminer le genre des visages. Par l'intermédiaire d'un enregistrement de l'activité en IRMf, ils ont montré que les HFS, en comparaison aux BFS, étaient associées à une activité accrue du gyrus fusiforme. L'amygdale était, de manière générale, plus activée face à des visages apeurés que neutres. Néanmoins, cet effet n'était significatif que si l'image présentée était en BFS ou non filtrée (donc quand l'information issue des BFS était disponible). Des réponses similaires ont été observées au niveau du CS et du thalamus. Pour les auteurs, ces résultats soulignent une spécificité de la voie sous-corticale aux BFS. La voie corticale serait quant à elle plus sensible aux HFS. Plus récemment, une étude avec un enregistrement intracrânien de l'activité de l'amygdale est venue appuyer cette hypothèse. Ainsi, dans une tâche similaire à celle de Vuilleumier et al. (2003), Méndez-Bértolo et al. (2016) ont observé des activations plus fortes pour des visages neutres que des visages apeurés dès 74 ms après l'apparition du stimulus. Cet effet n'était présent que lorsque les images étaient affichées en BFS, ou non filtrées. La Figure 1.17 présente un aperçu des stimuli utilisés, ainsi que des résultats obtenus dans les études de Vuilleumier et al. (2003) et Méndez-Bértolo et al. (2016). Certaines études en EEG ont également reporté une activité précoce, plus forte pour des visages émotionnels que neutres, limitée aux BFS (Nakashima et al., 2008 ; Pourtois et al., 2005 ; Vlamings et al., 2009 ; un effet qui n'est néanmoins pas toujours répliqué, voir par exemple Holmes, Winston et al., 2005 ; Jessen et Grossmann, 2017). Pour finir, une étude récente chez un patient atteint de cécité corticale a souligné que la réponse de l'amygdale induite par la présentation de visages apeurés dans le champ visuel aveugle n'apparaissait plus lorsque les images étaient présentées en HFS (Burra et al., 2019).

Au niveau comportemental, bien qu'il semble y avoir un usage flexible des fréquences spatiales pour le décodage des expressions, certaines études ont aussi rapporté un statut particulier des BFS dans la capture de l'attention par les visages émotionnels. Par exemple, dans une étude de Bannerman et al. (2012a) des participants devaient faire une saccade vers un visage qui apparaissait en périphérie, non filtré, en BFS, ou en HFS. En HFS, il n'y avait pas de différence entre les expressions. En BFS, au contraire, les saccades étaient effectuées plus rapidement pour aller vers des visages apeurés que joyeux. Dans une autre étude, Holmes et al. (2005) ont présenté une série d'expériences, dans lesquelles un visage apeuré et un visage neutre étaient présentés brièvement, en HFS ou en BFS. Ils étaient suivis par l'apparition d'une barre sur l'emplacement de l'un des visages, dont les participants devaient catégoriser l'orientation. L'identification de l'orientation des barres était effectuée plus rapidement lorsqu'elle remplaçait un visage apeuré que neutre, mais seulement en BFS (pour une méthode et des résultats similaires, voir Bocanegra et Zeelenberg, 2009).

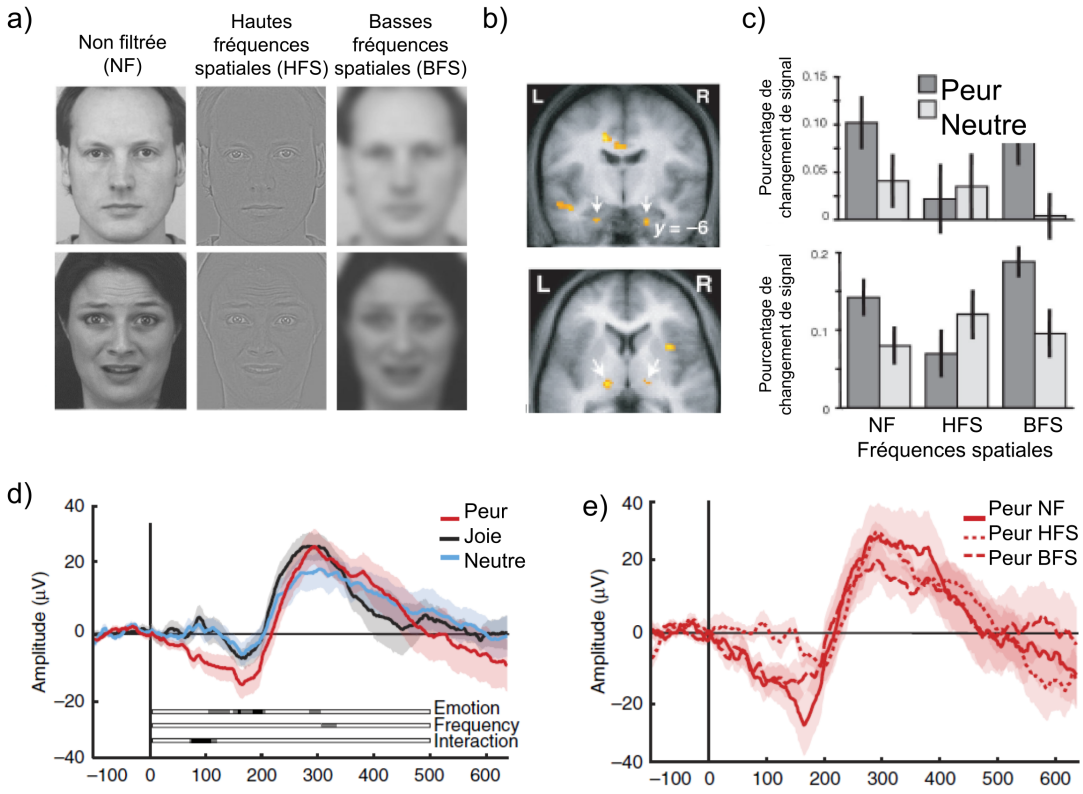

Figure 1.17 – Fréquences spatiales et voie sous-corticale. **(a)** Exemple de stimuli utilisés dans l'étude de Vuilleumier et al. (2003). Visages neutres (haut) ou apeurés (bas), non filtrés (gauche), en HFS (milieu) ou en BFS (droite). **(b)** Réponse de l'amygdale (haut) et du thalamus (bas) face à des visages apeurés ou neutres dans l'étude de Vuilleumier et al. (2003). La région du thalamus s'étend vers le CS. **(c)** Activité moyenne de l'amygdale (haut) et du thalamus (bas) en fonction des différentes conditions expérimentales dans l'étude de Vuilleumier et al. (2003). **(d)** Activité moyenne de l'amygdale en fonction des émotions dans l'étude de Méndez-Bértolo et al. (2016). **(e)** Activité moyenne de l'amygdale en fonction des fréquences spatiales pour les visages apeurés dans l'étude de Méndez-Bértolo et al. (2016). À travers ces résultats, nous pouvons voir que la réponse plus forte de l'amygdale face aux visages apeurés que neutres est spécifique aux images qui contiennent des BFS.

1.8.2 ... Mais aussi des remises en questions

Bien que les études qui argumentent en faveur de la voie sous-corticale telle que nous l'avons définie précédemment ne manquent pas, plusieurs études ont remis en question son existence, ou ont remis en cause certaines de ses caractéristiques. Par exemple, en 2010, Pessoa et Adolphs ont établi une revue des données qui remettent en question l'idée qu'une telle voie joue un rôle prépondérant dans le traitement des stimuli visuels émotionnels chez l'homme (pour une réponse à leurs critiques, voir de Gelder et al., 2011). Ils proposent un modèle dit multi-vagues dans lequel l'amygdale aurait pour rôle de coordonner les réseaux corticaux pendant l'évaluation de la pertinence biologique des stimuli visuels. Dans ce cadre, le cortex aurait un rôle plus important que ce qui lui était traditionnellement attribué. Cette remise en question est notamment appuyée par certaines études qui ont

mis en évidence des réponses dans le cortex visuel qui se produisaient avec des latences similaires à celles observées dans les zones sous-corticales (par exemple Andino et al., 2009 ; Krolak-Salmon et al., 2004 ; Ouellette et Casanova, 2006 ; Schmolesky et al., 1998). De plus, une étude chez un patient ayant une lésion complète de l'amygdale n'a pas montré de répercussions sur le traitement des visages apeurés (Tsuchiya et al., 2009).

Aussi, sans remettre en question son existence, plusieurs auteurs ont critiqué la sensibilité de cette voie sous-corticale aux expressions ou aux fréquences spatiales. Par exemple, dans une étude menée par Ottaviani et al. (2012) aucun effet des émotions n'a été observé sur l'activité de l'amygdale lorsque les images étaient présentées en BFS. Ensuite, Corradi-Dell'Acqua et al. (2014) ont analysé l'activité en IRMf en réponse à des images hybrides contenant les HFS d'un visage émotionnel et les BFS d'un visage neutre ou inversement. Chez les participants contrôles, ils n'ont pas observé de différence entre les visages émotionnels et neutres en BFS (cette différence était néanmoins observée en HFS). Certaines études ont utilisé la modélisation causale dynamique (ou *Dynamic Causal Modeling* ; DCM) pour évaluer la connectivité entre les régions lors de la perception de visages exprimant différentes émotions (Garvert et al., 2014 ; McFadyen et al., 2017). Plus précisément, cette méthode va calculer la probabilité de différents modèles (définis par les expérimentateurs en fonction du cadre théorique et des hypothèses), en se basant sur des données neurophysiologiques (ici, issues d'enregistrements de l'activité en magnétoencéphalographie ; MEG). Dans ces études les auteurs avaient comparé, entre autres, des modèles comportant seulement une voie corticale (reliant le CGL, V1 et l'amygdale), et des modèles comportant à la fois une voie corticale et une voie sous-corticale (reliant le pulvinar et l'amygdale). D'après leurs résultats, les modèles incluant la voie sous-corticale sont les plus probables. Cependant, cette voie ne serait influencée ni par les émotions du visage ni par les fréquences spatiales.

Pour finir, sans remettre en question l'idée que cette voie existe et puisse être modulée par les expressions faciales certains modèles suggèrent qu'elle aurait pour fonction de détecter des visages. Dans le cadre d'une revue de la littérature, Johnson a proposé un modèle en double voie du traitement des visages : une voie sous-corticale pour la détection et une voie corticale pour l'identification (Johnson, 2005 ; Johnson et al., 2015). La voie sous-corticale serait rapide, opérerait sur la base d'une information en BFS et impliquerait le CS, le pulvinar et l'amygdale. Pour Johnson, cette voie expliquerait les résultats d'études chez les nourrissons, qui montrent que certaines informations sur les caractéristiques des visages sont disponibles dès la naissance. En effet, les nouveau-nés humains s'orientent préférentiellement vers des modèles schématiques simples ressemblant à des visages (Johnson et al., 1991). La voie sous-corticale opérerait en parallèle au traitement cortical des visages et pourrait moduler son activité. Ainsi, certaines études ont montré que le degré d'activation des structures sous-corticales peut prédire l'activation des aires corticales sélectives aux visages (George et al., 2001 ; Morris et al., 1998). Les différences d'activations en fonction des expressions pourraient s'expliquer par la proximité avec le modèle schématique de visage. Cependant, la voie sous-corticale n'aurait pas pour fonction de discriminer les expressions.

1.8.3 L'égalisation du contraste, un biais méthodologique ?

Ainsi, la littérature concernant les effets des fréquences spatiales et des émotions sur le comportement ou sur les activations cérébrales est assez hétérogène. Dans l'étude de McFadyen et al. (2017), les auteurs suggèrent que l'absence de sensibilité aux fréquences spatiales et aux émotions sur la voie sous-corticale dans leur étude pourrait s'expliquer par le fait qu'ils utilisent des images égalisées en contraste de luminance. En fait, naturellement, après un filtrage des fréquences spatiales, les BFS et les HFS diffèrent, non seulement en termes de contenu fréquentiel, mais aussi en termes de contraste de luminance (que nous appellerons dans la suite de ce manuscrit contraste). Cela vient du fait que, dans le spectre d'amplitude des images, l'énergie décroît avec l'augmentation des fréquences spatiales (cette décroissance peut être approximée par une fonction en $1/f^{\alpha}$, avec f qui correspond à la fréquence spatiale, et α compris entre 0,8 et 1,5 ; Field, 1987 ; Loftus et Harley, 2005 ; Tolhurst et al., 1992 ; Van der Schaaf et van Hateren, 1996). Ainsi, dans les études où le contraste n'est pas égalisé après le filtrage, le contraste est plus important dans les BFS que dans les HFS. Donc, lors de la comparaison de données comportementales ou neurophysiologiques, le contraste pourrait aussi bien expliquer les différences observées entre les BFS et les HFS, car la comparaison entre les BFS et les HFS reviendrait à comparer des images faiblement contrastées à des images plus fortement contrastées.

Ainsi, le contraste élevé des BFS pourrait favoriser leur traitement, car les stimuli plus contrastés sont généralement mieux détectés par le système visuel. Par exemple, l'augmentation du contraste favorise la détection de réseaux sinusoïdaux (c'est-à-dire que les temps de réaction sont plus faibles avec un fort contraste ; Lupp et al., 1976 ; Vassilev et Mitov, 1976). Dans une étude récente, Perfetto et al. (2020) ont directement testé les effets de l'égalisation du contraste sur la catégorisation rapide de scènes filtrées (plus précisément, ils ont utilisé des scènes de plages, de forêts, de montagnes, d'autoroutes, de rues ou de bureaux). Ils ont observé de meilleures performances en BFS qu'en HFS, mais seulement quand les images n'étaient pas égalisées en contraste (voir Figure 1.18). Ce résultat suggère que la différence entre les BFS et les HFS peut s'expliquer par la différence de contraste. Dans le cadre de la catégorisation d'expressions faciales (neutres ou apeurées), Vlamings et al. (2009) ont également observé que les participants étaient plus rapides en BFS qu'en HFS, mais cet effet était réduit lorsque le contraste était égalisé. Généralement, il semble que la normalisation du contraste induise une augmentation des performances de catégorisation des images en HFS. Au niveau des activations cérébrales, le contraste pourrait aussi avoir un effet positif sur l'activation de plusieurs zones impliquées dans la perception visuelle, notamment V1 et l'amygdale (Boynton, 2005 ; Boynton et al., 1996 ; Goodyear et Menon, 1998 ; Inagaki et al., 2012). Dans le contexte de la perception de scènes, Kauffmann, Ramanoël et al. (2015) ont observé des activations plus fortes dans l'aire parahippocampique des lieux pour les images en BFS (ou non filtrées) qu'en HFS, mais cet effet était inversé lorsque les images étaient égalisées en contraste. Dans l'étude de McFadyen et al. (2017), il était précisé que le contraste était égalisé, ce qui n'était pas le cas dans d'autres études qui ont rapporté un effet des fréquences spatiales et des émotions (Méndez-Bértolo et al., 2016 ; Vuilleumier et al., 2003).

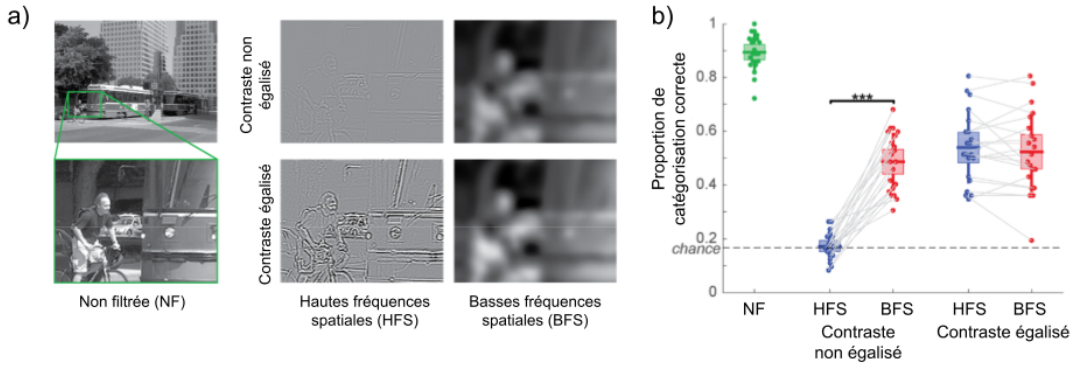

Figure 1.18 – Effets de l'égalisation du contraste dans le cadre de la catégorisation de scènes. **(a)** Stimuli et **(b)** résultats de l'étude de Perfetto et al. (2020). Les scènes étaient présentées non filtrées, en HFS ou en BFS, avec ou sans égalisation du contraste. On peut voir que la catégorisation est meilleure en BFS qu'en HFS mais seulement quand le contraste n'est pas égalisé.

1.8.4 Lien entre la voie sous-corticale et la programmation de saccades dans le cadre du traitement des expressions faciales

Dans les paragraphes précédents, nous avons mis en évidence un réseau neural qui pourrait sous-tendre l'orientation de l'attention vers les visages émotionnels : le réseau sous-cortical. Au centre de ce réseau, l'amygdale serait la région qui évaluerait la pertinence des stimuli visuels, et qui pourrait ainsi attribuer un poids plus important aux visages émotionnels (en particulier apeurés) plutôt que neutres. Cette évaluation rapide de la pertinence va permettre l'élaboration de réponses comportementales adaptées, par exemple la génération d'une saccade vers un stimulus pertinent. Dans cette dernière section, et pour conclure cette introduction théorique, nous allons nous intéresser aux mécanismes impliqués dans l'intégration de l'information émotionnelle dans le système oculomoteur. Ces mécanismes ont fait l'objet d'une récente revue de la littérature, dans laquelle Mulckhuyse (2018) souligne que, s'interroger sur l'intégration de l'information émotionnelle dans la programmation des mouvements oculaires, revient à se demander comment l'information est transmise de l'amygdale (impliquée dans l'évaluation de la pertinence émotionnelle) au CS (impliqué dans la phase finale de la programmation des mouvements oculaires). Dans cette revue, trois possibilités ont été mises en avant pour expliquer l'intégration de l'information émotionnelle dans le système oculomoteur. Ces possibilités sont représentées sur la Figure 1.19. Une première possibilité impliquerait la voie sous-corticale définie précédemment, qui relie le CS, le pulvinar et l'amygdale (Figure 1.19, lignes violettes). Les données concernant la direction des connexions entre ces régions sont peu nombreuses, mais il faudrait que les connexions qui relient le CS, le pulvinar et l'amygdale soient bilatérales pour que l'information puisse revenir rapidement par cette voie. Une seconde possibilité impliquerait la voie visuelle corticale. L'activité de l'amygdale amplifierait le traitement sensoriel des stimuli émotionnels dans le cortex visuel, qui projetterait ensuite l'information vers le CS (Figure 1.19, lignes vertes foncées). Une troisième possibilité, qui ne sera pas plus amplement développée dans ce travail de

thèse, impliquerait des projections corticales de l'amygdale aux zones frontales, telles que l'OFC, qui relaieraient ensuite l'information vers le CS (Figure 1.19, lignes vertes claires).

Ce modèle constitue un résumé simplifié des interactions entre le système oculomoteur et l'amygdale dans lequel seule l'amygdale est considérée pour l'évaluation de la saillance émotionnelle. Néanmoins, d'autres régions pourraient remplir un rôle similaire et une multitude d'autres voies pourraient être impliquées, opérant en parallèle ou non (Pessoa et Adolphs, 2010 ; Vuilleumier, 2015).

1.8 Arguments en faveur de la voie sous-corticale et débats actuels - Points clés

- Les patients atteints de cécité corticale sont capables de répondre à des expressions faciales présentées dans leur champ visuel aveugle.
- Des études en neuroimagie mettent en avant une discrimination des visages neutres et apeurés dans l'amygdale basée sur les BFS.
- Certains auteurs remettent en question le modèle en double voie pour le traitement des expressions faciales. Ils donnent par exemple plus d'importance au traitement cortical, ou remettent en question l'implication de la voie sous-corticale dans la discrimination des émotions.
- L'égalisation (ou non) du contraste de luminance pourrait avoir un impact sur les réponses à des visages émotionnels présentés sous différentes conditions de fréquences spatiales.

1.9 Problématique de la thèse et organisation du manuscrit

L'objectif des travaux menés dans le cadre de cette thèse était de préciser comment et quand est-ce que les expressions faciales émotionnelles vont capturer l'attention, et par extension les mouvements oculaires, en comparaison a des expressions faciales neutres. Comme nous l'avons vu, plusieurs hypothèses sont envisageables concernant l'intégration de l'information émotionnelle dans la programmation de mouvements oculaires, qui pourrait arriver à différents moments dans le décours temporel de la perception visuelle. Nous testerons particulièrement l'hypothèse d'une modulation rapide (< 100 ms), basée sur le traitement des BFS par la voie sous-corticale, qui favoriserait l'orientation de l'attention vers des visages émotionnels, particulièrement apeurés. La suite de ce manuscrit est divisée en trois chapitres expérimentaux (chapitres 2, 3, 4) et un chapitre de discussion (chapitre 5).

Le **chapitre 2** se présente sous la forme d'un article, publié dans la revue *Cognitive Science* en octobre 2021. Cet article regroupe deux expériences comportementales en choix saccadique, et une simulation par un réseau de neurones artificiel. La première expérience avait pour but d'étudier la capture de l'attention par des expressions joyeuses ou apeurées à un niveau précoce du traitement de l'information visuelle, au niveau de la détection d'un visage (c'est-à-dire dès 100-120 ms après la présentation des stimuli). La seconde expérience avait pour but d'évaluer la détection d'un visage émotionnel (joyeux ou apeuré)

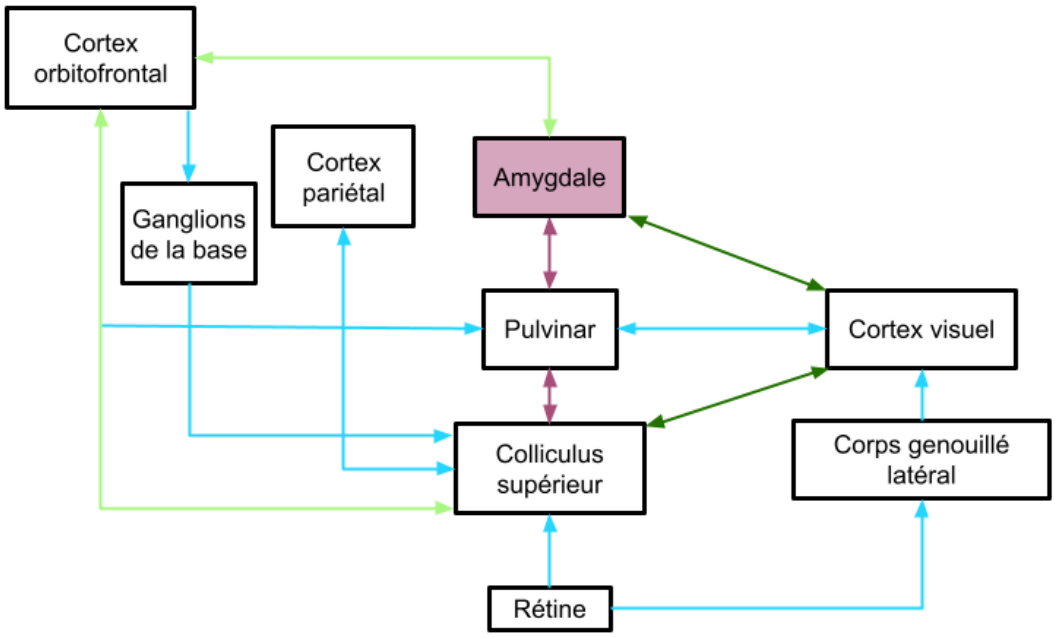

Figure 1.19 – Modèle simplifié du système oculomoteur comprenant l'amygdale, et trois voies possibles par lesquelles l'amygdale pourrait moduler le comportement oculomoteur. En violet, une boucle correspondant à la voie sous-corticale qui relie l'amygdale au CS par l'intermédiaire du pulvinar. En vert clair et vert foncé, les connexions corticales par lesquelles l'amygdale pourrait amplifier le traitement sensoriel dans les zones visuelles (vert foncé), ou par lesquelles l'amygdale pourrait améliorer le traitement des caractéristiques visuelles grossières (vert clair). Figure adaptée de Mulckhuyse (2018).

par rapport à un visage neutre. Les résultats de cette seconde expérience ont été comparés aux résultats d'un réseau de neurones artificiel qui effectuait une tâche similaire, afin de dissocier les contributions physiques et émotionnelles. Ce chapitre comporte en plus une expérience complémentaire qui s'intéressait à l'influence des expressions faciales dans la détection du genre. De manière générale, **ce chapitre a pour but de comparer l'influence des expressions faciales dans différentes tâches, à différents moments dans le décours temporel du traitement de l'information visuelle**.

Le **chapitre 3** est composé de deux articles soumis. Le premier article présente une expérience en choix saccadique, similaire à la seconde expérience de l'article du chapitre 2. L'expérience comportementale avait pour but d'évaluer la détection d'un visage émotionnel (joyeux ou apeuré) par rapport à un visage neutre, en fonction de différentes conditions de filtrage, et de l'égalisation du contraste. Il présente également une analyse de la saillance des stimuli. Nous avions pour objectif d'évaluer la relation entre les performances des participants et la saillance des régions diagnostiques au décodage des expressions du visage, en particulier les yeux et la bouche. Le second article présente un réseau de neurones artificiel qui permet de mettre en avant les régions utiles à la discrimination d'un visage émotionnel et d'un visage neutre. De manière générale, **ce chapitre a pour**

but d'évaluer l'effet des fréquences spatiales et de l'égalisation du contraste dans la détection de visages émotionnels, ainsi que l'importance de différentes parties du visage.

Le **chapitre 4** décrit une expérience en IRMf. Plus spécifiquement, nous avons mesuré l'activité cérébrale en réponse à des visages apeurés ou neutres dans plusieurs régions d'intérêt (notamment l'amygdale, le pulvinar, le CS, la FFA et l'OFA). Les visages étaient présentés en HFS ou en BFS, avec ou sans égalisation du contraste, et les participants devaient les catégoriser selon leur genre. **Le but de chapitre était de tester les réponses neurales de régions cruciales dans le cadre d'un modèle en double voie en fonction des expressions faciales, des fréquences spatiales et de l'égalisation du contraste.** Le Tableau 1.1 fait le bilan des expériences présentées dans ce travail de thèse, ainsi que de leur répartition au sein des chapitres expérimentaux.

Le **chapitre 5** présente une discussion générale de l'ensemble des résultats des chapitres expérimentaux.

Chapitres	Expériences	Méthode	Publication
Chapitre 2	**Expérience 1** - Visage vs Véhicule	Choix saccadique	Article 1, publié
	Expérience 2 - Visage émotionnel vs Visage neutre	Choix saccadique	Article 1, publié
	Simulation de l'Expérience 2	MLP	Article 1, publié
	Expérience 3 - Visage masculin vs visage féminin	Choix saccadique	
Chapitre 3	**Expérience 4** - Visage émotionnel vs Visage neutre - Effet des fréquences spatiales et de l'égalisation du contraste de luminance	Choix saccadique	Article 2, soumis
	Analyse de saillance		Article 2, soumis
	Simulation de l'Expérience 4	CNN	Article 3, soumis
Chapitre 4	**Expérience 5** - Bases cérébrales du traitement des expressions faciales et lien avec les fréquences spatiales et l'égalisation du contraste de luminance	IRMf	

Table 1.1 – Organisation des chapitres expérimentaux.

Chapitre 2

Rôle de la tâche et décours temporel de la perception des visages émotionnels

Table des matières

2.1 Préface

Comme nous l'avons vu dans le chapitre précédent, l'aspect automatique de la capture de l'attention et du regard par les visages émotionnels est débattu. Si certaines études suggèrent que les visages émotionnels puissent être différenciés très tôt, dès 100 ms à partir de l'activité cérébrale (par exemple Méndez-Bértolo et al., 2016 ; W. Sato et al., 2011), les résultats comportementaux sont plus mitigés. En particulier, les études utilisant des paradigmes de recherche visuelle avec un enregistrement des mouvements oculaires ne mettent pas toutes en avant une capture du regard plus importante pour les visages émotionnels que neutres (par exemple Devue et Grimshaw, 2017 ; Hunt et al., 2007 ; Mulckhuyse, 2018). L'objectif principal de ce chapitre est de tester l'hypothèse d'une capture du regard par les visages émotionnels (en comparaison à des visages neutres) qui interviendrait rapidement et indépendamment de la tâche. Dans ce chapitre, nous présentons une série de trois expériences en choix saccadique. Toutes ces expériences incluent des visages émotionnels (joyeux ou apeurés) et des visages neutres. Mais, elles impliquent des consignes différentes et des réponses à différentes échelles temporelles. Plus précisément, dans la première expérience (Expérience 1), des visages étaient opposés à des véhicules, et les participants devaient faire une saccade le plus rapidement possible vers le visage ou le véhicule. Dans la seconde expérience (Expérience 2), des visages émotionnels étaient opposés à des visages neutres et les participants devaient faire une saccade le plus rapidement possible vers le visage émotionnel ou neutre. Dans la troisième expérience (Expérience 3), de la même manière que dans l'Expérience 2, des visages émotionnels étaient opposés à des visages neutres. Mais, cette fois, les participants devaient faire une

saccade vers le visage masculin ou féminin [1]. L'Expérience 1 est connue pour générer des réponses très rapides vers les visages. Nous nous attendons à ce que les réponses soient un peu plus tardives dans les Expériences 2 et 3. Dans les Expériences 1 et 3, nous avons testé l'effet du traitement implicite des expressions faciales, car elles n'étaient pas pertinentes pour la tâche des participants. Au contraire, dans l'Expérience 2, c'est l'effet du traitement explicite des expressions faciales qui est testé, car on demande directement aux participants de les prendre en compte. Si la capture du regard par les visages émotionnels, en particulier apeurés, est automatique, elle devrait intervenir rapidement et indépendamment de la tâche des participants.

Dans les trois expériences, nous avons analysé la première saccade des participants. Si cette saccade est dirigée vers la cible, elle est considérée comme une saccade correcte, sinon elle est considérée comme une saccade erreur. Trois variables d'intérêt ont été analysées dans différentes conditions expérimentales. D'abord, la proportion de saccades correctes, qui correspond au nombre de saccades correctes divisé par le nombre d'essais. Ensuite, la latence des saccades correctes, qui correspond au temps qui sépare l'apparition des images et le déclenchement d'une saccade correcte. Ces deux variables nous permettent d'évaluer l'intensité de la capture du regard par les visages émotionnels, en comparaison aux visages neutres. En effet, si les visages émotionnels capturent le regard plus que les visages neutres, ils devraient être associés à des latences plus faibles et des proportions de saccades correctes plus élevées que les visages neutres lorsqu'ils sont la cible des participants. Pour finir, nous avons aussi analysé les points d'arrivée des saccades correctes, afin d'évaluer la distribution de l'attention dans le visage pendant la programmation des saccades. Plus précisément, nous avons analysé la distance verticale entre le point d'arrivée de la saccade et le centre de l'image (qui correspond à la position des yeux). Il est bien connu qu'en fonction de l'émotion les régions diagnostiques sont différentes. En préparant un mouvement oculaire vers un visage, nous supposons que l'attention est dirigée vers celui-ci, puisqu'il sera la cible de la prochaine saccade. Cependant, nous nous attendions à ce que les caractéristiques diagnostiques ou saillantes de chaque expression soient capables de moduler l'allocation de l'attention au sein du visage en attirant les points d'arrivée des saccades vers elles.

Dans ce chapitre, nous présenterons d'abord les résultats obtenus pour les deux premières expériences, qui ont fait l'objet d'une publication. Ces expériences sont accompagnées dans l'article d'une simulation, dont le but était d'expliquer les résultats obtenus dans l'Expérience 2. L'Expérience 3, qui n'a pas fait l'objet d'une publication, est présentée ensuite.

2.2 Article 1

Dans ce premier article, nous nous sommes intéressés à l'impact des expressions faciales émotionnelles sur la programmation de saccades. Dans une première expérience (Visage vs Véhicule), l'objectif principal était de déterminer si la présence d'une émotion peut faciliter le déclenchement rapide des saccades vers les visages. Nous avons reproduit une

1. Dans les Expériences 2 et 3, à chaque essai, un visage féminin était opposé à un visage masculin.

tâche de choix saccadique opposant des visages et des véhicules, en utilisant des visages qui présentaient une expression joyeuse, apeurée ou neutre. Les participants devaient faire une saccade le plus rapidement possible vers le véhicule (dans une session) ou vers le visage (dans une autre session). Conformément aux résultats de la littérature, nous nous attendions à de meilleures performances (c'est-à-dire à des proportions de saccades correctes plus élevées et des latences plus courtes) lorsque la cible était un visage plutôt qu'un véhicule. Nous nous attendions aussi à ce que les visages émotionnels attirent plus facilement le regard (c'est-à-dire à ce qu'il soient associés à des proportions de saccades correctes plus élevées et des latences plus courtes) lorsqu'ils sont la cible des participants que les visages neutres, en raison de leur pertinence émotionnelle. En ce qui concerne les différences entre les visages joyeux et apeurés, nous avons formulé deux hypothèses alternatives. Premièrement, en lien avec l'hypothèse d'une voie sous-corticale impliquée dans la perception rapide des stimuli menaçants, notamment les visages exprimant de la peur, nous devrions observer de meilleures performances lorsque la cible est un visage apeuré plutôt qu'un visage joyeux. Inversement, en supposant que les visages joyeux sont plus faciles à détecter en raison de leurs caractéristiques physiques, nous devrions observer de meilleures performances lorsque la cible est un visage joyeux qu'un visage apeuré. En considérant que les points d'arrivée des saccades reflètent la distribution de l'attention dans le visage, nous nous attendions à ce qu'ils soient plus proches des yeux pour les visages apeurés que pour les visages joyeux (du fait que la région de la bouche est plus importance et plus saillante dans les visages joyeux qu'apeurés).

Les résultats de cette première expérience ont montré que les participants faisaient moins d'erreurs, et ont exécuté leurs saccades plus rapidement lorsque la cible était un visage plutôt qu'un véhicule (proportions de saccades correctes moyennes : 0.88 pour les visages et 0.86 pour les véhicules ; latences moyennes : 176 ms pour les visages et 191 ms pour les véhicules). Concernant les expressions faciales, aucun effet n'a été observé sur les latences ou les proportions de saccades correctes. Néanmoins, un effet des expressions faciales a été observé sur les points d'arrivée des saccades. En effet, les saccades atterrissaient généralement autour des yeux ; cependant les points d'arrivée des saccades étaient plus haut dans le visage lorsque le visage était neutre que lorsqu'il était apeuré ou joyeux, et lorsque le visage était apeuré plutôt que joyeux. En conclusion, les visages émotionnels n'ont pas attiré le regard plus que les visages neutres. Nous supposons que les émotions ne sont pas encore décodées à des temps aussi faibles, avant que le visage ne soit détecté. Nous supposons néanmoins que les caractéristiques saillantes des visages, comme la bouche ou les yeux, peuvent moduler la distribution de l'attention dans le visage et la programmation des saccades, même à ce stade précoce. Cela se traduirait par un décalage des points d'arrivée vers les caractéristiques les plus saillantes.

Dans une seconde expérience (Visage émotionnel vs Visage neutre), nous avons toujours utilisé un paradigme de choix saccadique, mais cette fois en opposant un visage émotionnel (joyeux ou apeuré) et un visage neutre, afin de tester directement la détection des expressions faciales. Les hypothèses testées dans cette expérience étaient les mêmes que celles testées dans l'Expérience 1. Nous nous attendions à observer de meilleures

performances (c'est-à-dire des proportions de saccades correctes plus élevées et des latences plus courtes) lorsque la cible était un visage émotionnel que neutre. En ce qui concerne les différences entre les visages joyeux et apeurés, nous avons formulé deux hypothèses alternatives. En lien avec l'hypothèse d'une voie sous-corticale impliquée dans la perception rapide des stimuli liés à une menace, nous devrions observer de meilleures performances lorsque la cible est un visage apeuré. Inversement, en supposant que les visages joyeux sont plus faciles à détecter en raison de leurs caractéristiques physiques, nous devrions observer de meilleures performances lorsque la cible est un visage joyeux. Nous nous attendions toujours à ce que les points d'arrivée des saccades soient plus proches des yeux pour les visages apeurés que pour les visages joyeux.

Les résultats de cette seconde expérience ont montré que les participants faisaient moins d'erreurs, et ont exécuté les saccades plus rapidement lorsque la cible était un visage émotionnel qu'un visage neutre (proportions de saccades correctes moyennes : 0.68 ms pour les visages émotionnels et 0.62 ms pour les visages neutres ; latences moyennes : 249 ms pour les visages émotionnels et 277 ms pour les visages neutres). De plus, lorsque la cible était un visage émotionnel, les participants faisaient moins d'erreurs avec un visage joyeux qu'apeuré. Ainsi les visages émotionnels peuvent attirer l'attention plus que les visages neutres ; cet effet semble cependant dépendant de la tâche car il n'est observé que dans l'Expérience 2. En ce qui concerne les points d'arrivée des saccades, ils étaient encore une fois situés plus haut lorsque la cible était un visage apeuré plutôt qu'un visage joyeux. Par rapport à la première expérience (dans laquelle les saccades avaient tendance à se poser autour des yeux), les saccades avaient tendance à se poser autour du nez dans l'Expérience 2. Cela suggère que, même si les caractéristiques locales sont capables de capter l'attention d'une manière automatique, le poids alloué à ces caractéristiques peut être modifié par la tâche.

La simulation mise en place avait pour objectif de mieux comprendre l'effet obtenu dans l'Expérience 2 concernant la meilleure détection de visages émotionnels joyeux. Nous supposons qu'une hypothèse liée aux propriétés physiques des stimuli est capable d'expliquer l'avantage des visages joyeux, sans que les processus émotionnels interviennent. Nous avons utilisé un réseau de neurones artificiel simple, de type perceptron multicouche (ou *multilayer perceptron* ; MLP), pour tester cette hypothèse. Le but était de quantifier les différences physiques entre les visages neutres et les visages apeurés ou joyeux. Nous n'avons pas simulé directement de réponses saccadiques. Cependant, la tâche du réseau était similaire, puisqu'il s'agissait de trouver l'emplacement d'un visage émotionnel par rapport à un visage neutre. Plus précisément, le réseau a été entraîné et testé sur sa capacité à discriminer des paires de visages de type émotionnel-neutre et des paires de visages de type neutre-émotionnel (le visage émotionnel se trouvant soit à droite, soit à gauche de la paire). Les paires de visages présentées au réseau étaient les mêmes que celles présentées aux participants dans l'Expérience 2. Pour chaque paire testée, nous enregistrions la réponse du réseau qui pouvait être correcte (si le réseau avait bien catégorisé la paire) ou non, et la proportion de saccades correctes était analysée. Nous nous attendions à ce que les performances du réseau soient meilleures avec un visage émotionnel joyeux

plutôt qu'apeuré. En effet, si les visages joyeux et neutres sont plus faciles à distinguer que les visages apeurés et neutres sur la base des statistiques des images, cela pourrait expliquer l'avantage des visages joyeux obtenu dans l'Expérience 2. Conformément à notre hypothèse, le réseau de neurones a montré de meilleures performances de catégorisation lorsque le visage émotionnel était joyeux plutôt qu'apeuré. Cela suppose que l'avantage des visages joyeux dans une tâche de détection des visages émotionnels peut s'expliquer par les statistiques des images. Notons que dans cet article ainsi que dans les articles suivants, les termes *perceptual factors* et *perceptual saliency* sont utilisés pour référer aux facteurs physiques, indépendants de l'observateur et du système perceptif.

L'Article 1 a été publié dans la revue *Cognitive Science* en octobre 2021. Les annexes de cet article sont présentées à la fin du manuscrit de thèse, dans l'Appendice A. Le Tableau 2.1 dresse la liste des contributions de chaque auteur.

Contributeurs	Contributions
Léa Entzmann	Conception de l'expérience ; Recueil des données ; Analyse des données ; Simulation ; Rédaction du manuscrit ; Édition du manuscrit
Nathalie Guyader	Conception de l'expérience ; Édition du manuscrit
Louise Kauffmann	Conception de l'expérience ; Édition du manuscrit
Juliette Lenouvel	Recueil des données
Clémence Charles	Recueil des données
Carole Peyrin	Conception de l'expérience ; Édition du manuscrit
Roman Vuillaume	Conception de l'expérience (partie simulation) ; Édition du manuscrit (partie simulation)
Martial Mermillod	Conception de l'expérience ; Édition du manuscrit

Table 2.1 – Contributions des auteurs de l'Article 1.

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e13042

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13042

The Role of Emotional Content and Perceptual Saliency During the Programming of Saccades Toward Faces

Léa Entzmann,[a,b] Nathalie Guyader,[b] Louise Kauffmann,[a] Juliette Lenouvel,[a] Clémence Charles,[a] Carole Peyrin,[a] Roman Vuillaume,[c] Martial Mermillod[a]

[a]*LPNC, CNRS, Université Grenoble Alpes Université Savoie Mont Blanc*
[b]*GIPSA-lab, Université Grenoble Alpes CNRS Grenoble INP*
[c]*ImViA, Université Bourgogne Franche-Comté*

Received 30 October 2020; received in revised form 22 June 2021; accepted 10 August 2021

Abstract

Previous studies have shown that the human visual system can detect a face and elicit a saccadic eye movement toward it very efficiently compared to other categories of visual stimuli. In the first experiment, we tested the influence of facial expressions on fast face detection using a saccadic choice task. Face-vehicle pairs were simultaneously presented and participants were asked to saccade toward the target (the face or the vehicle). We observed that saccades toward faces were initiated faster, and more often in the correct direction, than saccades toward vehicles, regardless of the facial expressions (happy, fearful, or neutral). We also observed that saccade endpoints on face images were lower when the face was happy and higher when it was neutral. In the second experiment, we explicitly tested the detection of facial expressions. We used a saccadic choice task with emotional-neutral pairs of faces and participants were asked to saccade toward the emotional (happy or fearful) or the neutral face. Participants were faster when they were asked to saccade toward the emotional face. They also made fewer errors, especially when the emotional face was happy. Using computational modeling, we showed that this happy face advantage can, at least partly, be explained by perceptual factors. Also, saccade endpoints were lower when the target was happy than when it was fearful. Overall, we suggest that there is no automatic prioritization of emotional faces, at least for saccades with short latencies, but that salient local face features can automatically attract attention.

Keywords: Emotional facial expressions; Eye movements; Saccade programming; Neural computation; Time course

Correspondence should be sent to Léa Entzmann, Univ. Grenoble Alpes, 1251 Avenue Centrale, Bâtiment Michel Dubois, 38400 Saint-Martin-d'Hères, France. E-mail: lea.entzmann@univ.grenoble-alpes.fr

1. Introduction

The human visual system is extremely efficient in the rapid, preferential detection of socially relevant stimuli, such as faces. In particular, eye-tracking data have shown that when presented in visual scenes, faces immediately attract the gaze of observers, who then spend most of the exploration time looking at them (Cerf, Frady, & Koch, 2009; Coutrot & Guyader, 2014; Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010; Marat et al., 2009). Furthermore, saccades toward individual faces can be made continuously, at rates up to 6 faces/s (Martin, Davis, Riesenhuber, & Thorpe, 2018). Moreover, when presented along with a distractor image (e.g., a vehicle), face stimuli can be detected and elicit a saccade toward them very rapidly, while more time is needed for other objects (Crouzet, Kirchner, & Thorpe, 2010; Guyader, Chauvin, Boucart, & Peyrin, 2017; Kauffmann et al., 2019).

This last result has been highlighted in saccadic choice tasks, where saccades are used as behavioral responses. In such tasks, two images from different categories (e.g., a face and a vehicle) are displayed on the screen and participants are asked to make a saccade as fast as possible toward the image that contains the target category (i.e., the face or the vehicle). Such tasks have revealed that saccades toward faces can be reliably elicited in only 100 ms (Crouzet et al., 2010) and, overall, suggest that faces can attract the gaze very rapidly. This bias for faces during the saccadic choice task has been replicated across many studies and is robust to stimulus manipulations, such as grayscaling, thumbnail, and phase scrambling or spatial frequency filtering (Boucart et al., 2016, Guyader et al., 2017, Honey, Kirchner, & VanRullen, 2008, Crouzet & Thorpe, 2011, Kauffmann, Khazaz, Peyrin, & Guyader, 2021). Furthermore, it persists even if faces are opposed to distractors sharing a similar shape, degree of animacy, or structural homogeneity (Boucart et al., 2016, Kauffmann et al., 2021). In everyday life, humans are well accustomed to transmitting and decoding emotional information from faces by means of facial expressions used for social communication. The brain has consequently developed a number of specific and complex mechanisms, which are as yet not fully understood, to process emotionally relevant information from faces (for a review, see Adolphs, 2003). While face stimuli can guide attention very efficiently, it seems appropriate to ask how this can be modulated through facial expressions.

In fact, facial expressions are characterized by both a specific physical facial configuration and an emotion they are assumed to convey (Calvo & Nummenmaa, 2015). Generally, emotional stimuli have been found to be processed more efficiently than neutral ones. For example, an emotional object is more likely to be fixated first than a neutral object (Humphrey, Underwood, & Lambert, 2012; Niu, Todd, & Anderson, 2012). Also, many studies suggest that humans have evolved to preferentially orient their attention toward threatening stimuli. At the behavioral level, angry faces seem to be particularly well detected when presented among matrices of other expressions in visual search tasks (Fox & Damjanovic, 2006; Öhman, Lundqvist, & Esteves, 2001; Schubö, Gendolla, Meinecke, & Abele, 2006; Tipples, Atkinson, & Young, 2002; for a review, see Frischen, Eastwood, & Smilek, 2008). Also, neurophysiological data suggest that this attentional modulation is associated with activations in the limbic system, including the amygdala. In fact, classical models of emotional processes in the brain suppose that threat-related stimuli, in this case fearful faces in particular, can be detected

very rapidly through a subcortical pathway involving the superior colliculus, pulvinar, and amygdala. This pathway is thought to be magnocellular and to transmit coarse information in parallel to the finer and slower cortical processing (LeDoux, 2000; Morris, 1998; Öhman, 2005; Tamietto & de Gelder, 2010). This idea is supported by intracranial electroencephalography (EEG) recording in the amygdala showing activations as early as 74 ms for coarse fearful faces (Méndez-Bértolo et al., 2016).

However, the existence of such a pathway as well as its involvement in the detection of facial expressions remain a matter of debate, and it is possible that the subcortical route operates for faces irrespective of emotions (Fitzgerald, Angstadt, Jelsone, Nathan, & Phan, 2006; Garvert, Friston, Dolan, & Garrido, 2014; Johnson, 2005; McFadyen, Mermillod, Mattingley, Halász, & Garrido, 2017). Also, other behavioral studies, especially those using nonschematic faces, have found no prioritization of threatening faces (Becker, Anderson, Mortensen, Neufeld, & Neel, 2011; Calvo & Marrero, 2009; Calvo & Nummenmaa, 2011; Hunt, Cooper, Hungr, & Kingstone, 2007; Juth, Lundqvist, Karlsson, & Ohman, 2005; Lipp, Price, & Tellegen, 2009). For example, Calvo and Nummenmaa (2009) used a saccadic choice task with one emotional and one neutral face presented simultaneously and found that happy, rather than angry or fearful faces, were detected better. These observations are consistent with results from categorization task studies which used manual responses (Calvo & Lundqvist, 2008; Tottenham et al., 2009). In 2011, the same authors assessed the role of perceptual (e.g., luminance and local saliency) and semantic (e.g., affective valence) factors in the discrimination advantage of happy faces, and found only a contribution of mouth saliency. They suggested that this "happy face advantage" relies more on the saliency of perceptual features (like the open mouth) rather than the interpretation of the emotional content (Calvo & Nummenmaa, 2011). Indeed, the smiling mouth is more salient than any other region of happy and nonhappy faces (based on a combination of physical image properties, such as luminance, contrast, and spatial orientation; Calvo & Nummenmaa, 2008; Itti & Koch, 2000), suggesting that it can be used as a shortcut for the quick recognition of happy faces (Calvo & Nummenmaa, 2015). Thus, Horstmann, Lipp, and Becker (2012) found that visible teeth in angry or happy faces make them easier to detect in a crowd of neutral faces.

The first goal of this study is to determine whether the presence of an emotional expression can facilitate very fast face detection. Indeed, although many studies have focused on what drives the very fast saccades toward faces, these studies have mostly used neutral faces and it is still unclear whether emotional facial expression influences face detection. In the first experiment, we reproduced a saccadic choice task with face-vehicle pairs, using faces which portrayed either a happy, fearful, or neutral expression. We chose fearful rather than angry faces in order to establish a connection with neurophysiological data showing that such faces are processed very efficiently in the brain (LeDoux, 2000; Méndez-Bértolo et al., 2016). Contrary to previous studies, in which images were presented in natural contexts (Crouzet et al., 2010; Kauffmann et al., 2019; Kirchner & Thorpe, 2006), we used more prototypical stimuli. Indeed, classical databases of emotional faces are very prototypical, with all the faces being centered at the same position in the image. Hence, we also used very prototypical vehicle stimuli taken from the stimuli created by Kloth and used by Kloth, Itier, and Schweinberger (2013) in a study in which the authors tested neurophysiological responses to nonface objects

that are structurally similar to faces. In line with previous findings, we expected better performances (higher accuracy and shorter latency) for saccades toward face than saccades toward vehicle targets. Regardless of which specific expression can attract the attention the most, we generally expected emotional faces to attract the attention more than neutral faces because of their emotional relevance. Concerning the differences between happy and fearful faces, we formulated two alternative hypotheses. First, assuming that humans evolved to preferentially orient their attention toward threat, we should observe better performances toward fearful than happy face targets. Conversely, assuming that happy faces are easier to detect because of perceptual factors, we should find better performances toward happy than fearful face targets.

We also examined saccade endpoints in order to assess the distribution of attention within the face during saccade programming as a function of the expression. Indeed, saccade programming is thought to occur in a priority map in which both bottom-up (e.g., local saliencies) and top-down (e.g., emotional relevance) information is integrated. This priority map is assumed to follow a retinotopic organization associated with a winner-takes-all mechanism that guides the allocation of attention through orientation of the gaze (Belopolsky, 2015; Bisley & Mirpour, 2019; Fecteau & Munoz, 2006; Klink, Jentgens, & Lorteije, 2014; Theeuwes, 2019). Saccade endpoints are thus the result of a competition between multiple locations. It is well known that emotional facial expressions have different diagnostic features, mostly in the form of the eyes and the mouth (Eisenbarth & Alpers, 2011; Smith, Cottrell, Gosselin, & Schyns, 2005; Wegrzyn, Vogt, Kireclioglu, Schneider, & Kissler, 2017). For example, Smith et al. (2005) found that the recognition of happy and surprised faces is based more on the mouth, whereas the recognition of fearful and angry faces is based more on the eyes. While preparing an eye movement toward a face, we suppose that attention is directed toward it. However, we expected that the diagnostic or salient features of each expression would be able to modulate the allocation of attention within the face by attracting it. For example, if a happy face is presented, attention may be directed more specifically toward the mouth, which is more salient and diagnostic for this expression and, if a fearful face is presented, more attention may be directed toward the eyes. Thus, considering that saccade endpoints can reflect the distribution of attention within the face, we expected that they would be closer to the eyes for fearful than for happy faces.

In the first experiment, the influence of facial expressions was assessed at a very early stage of visual processing (with saccades that are often elicited in less than 200 ms) using a task that does not explicitly require the processing of facial expressions. Some studies suggest that expression decoding only occurs at a later stage of visual processing (after 180 ms; Kulke, 2019; Schyns, Petro, & Smith, 2009). It is, therefore, unclear whether expressions could have been decoded before saccade onset in the first experiment. In the second experiment, we explicitly assessed the detection of facial expressions in order to test the preferential processing of emotional faces displaying task-relevant expressions. In a way similar to Bannermann et al. (2009) and Calvo & Nummenmaa (2009), neutral and emotional faces (happy and fearful) were presented simultaneously in a saccadic choice task, and participants were asked to saccade toward the emotional or the neutral face. We still expected to find better performances when the target was an emotional than when it was a neutral face. Furthermore, if attention can be preferentially attracted by threat, we should observe better performances for fearful

faces. Alternatively, given that happy faces are easier to recognize, better performances might be expected for happy faces. However, even if this were the case, we still expected saccade endpoints to be closer to the eyes for fearful than for happy faces.

Finally, we present a computational model that simulates results from a task similar to the one presented in the second experiment. The goal of this simulation was to show that the happy face advantage that could be expected in the second experiment can be explained through perceptual factors, as suggested by Calvo & Nummenmaa (2011). Computational models are a useful tool for testing the role of the physical properties of inputs because they are not sensitive to high-level influences, such as the interpretation of the emotional content (Mermillod, Vermeulen, Lundqvist, & Niedenthal, 2009). If there is a happy face advantage and if it is mainly perceptual in nature, then the neural network should perform better with happy than with fearful faces.

2. Experiment 1: Face versus vehicle

2.1. Materials and methods

2.1.1. Participants

Sixty-seven participants were recruited at the local university to perform a saccadic choice task. Considering the simple nature of the task, six of them were removed from the statistical analysis due to their low proportion of correct responses (below 0.75 in both sessions), leaving a group of 61 participants for inclusion in the statistical analysis (29 females; $M \pm SD$: 21.86 $\pm$ 0.47 years; age range: 18–36 years). All of them had normal or corrected-to-normal visual acuity. Undergraduate psychology students received course credits for their participation in the experiment. All participants gave their informed written consent before the experiment, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.1.2. Stimuli

Stimuli consisted of 160 grayscale photographs, with 120 images containing a face (40 different faces with three emotions; fearful, happy, and neutral) and 40 containing a vehicle (Fig. 1). During the experiment, images were systematically presented in face-vehicle pairs. *Face stimuli* were chosen from the Karolinska Directed Emotional Faces database (KDEF; Lundqvist, Flykt, & Ohman, 1998), which is widely used in the field of emotion processing. *Vehicle stimuli* took the form of 40 pictures of cars taken from the stimuli used by Kloth et al. (2013). Each vehicle stimulus was duplicated three times to counterbalance the number of face and vehicle images (just as each of the 40 individual faces was seen three times for the three emotion conditions). The KDEF database is constructed in such a way that the vertical and horizontal positions of the eyes and mouth of each picture are set to the same positions on a digital grid. The original images have a size of 562 pixels x 762 pixels. However, for our experiment, they were cropped by 100 pixels at the top and bottom and were resized to 300 pixels x 300 pixels, corresponding to a coverage of 11 $\times$ 11 degrees of

Fig 1. Example of face and vehicle stimuli used in Experiment 1 (left), and their mean amplitude spectrum (right).

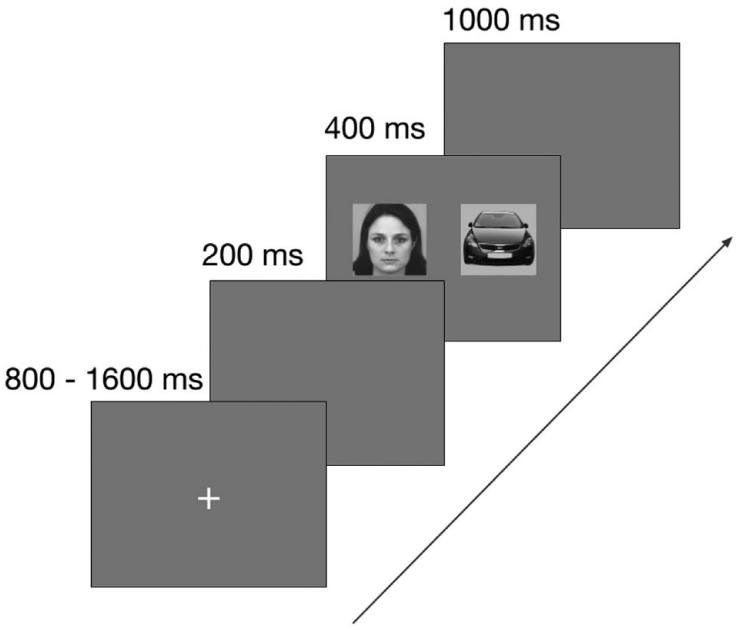

Fig 2. Time course of a trial for Experiment 1.

visual angle at a 57-cm viewing distance. This preprocessing allowed us to create a database of images with a size matching that was used for our previous papers (Guyader et al., 2017; Kauffmann et al., 2019), and to have images with close-up faces. Moreover, the position of the eyes corresponded to the middle of the image as shown in Fig. 3. After resizing, the images were equalized in terms of mean luminance and root mean squared contrast (mean luminance value of 127 and a mean contrast of 47, for pixel intensities between [0,255], based on mean luminance and contrast values of all the stimuli). Following this equalization step, all the images (faces and vehicles) globally had the same mean luminance and mean Root Mean

Square (RMS) contrast. A gamma correction was applied to each image to adapt the stimuli to the screen luminance. Vehicles were set against a gray background and were manually resized so that faces and vehicles had the same average spatial position and size. A training session consisting of six practice trials with 12 additional images not used in the experimental sessions (six vehicles and six faces) was performed at the beginning of the experiment to allow the participants to familiarize themselves with the task.

2.1.3. Procedure

Stimuli were displayed on a 24-inch screen with a spatial resolution of 1360 × 768 pixels and a refresh rate of 60 Hz. This resolution was used with a desktop computer. A keyboard was placed in front of the participants, allowing them to end the breaks by pressing the space-bar. The screen was raised a little at a distance of about 20 cm away from the keyboard and the back of the experimentation room was empty. Eye movements were recorded with an Eyelink 1000 (SR Research) eye-tracker with a 1000-Hz sampling frequency. Viewing was binocular, but only the position of the dominant eye was recorded. Saccades were automatically detected using the Eyelink software. Saccades were detected if they had a minimum velocity of 30 degrees/s, a minimum acceleration of 8000 degrees/s^2, and a minimum motion of 0.15 degrees. Blinks were detected when the pupil was partially or totally occluded, and fixations were detected when there was no blink and no saccade in progress. The experiment was divided into two sessions, with the order being counterbalanced between participants. In each session, each of the 120 different faces was displayed, once on the left and once on the right side of the screen, leading to 240 trials per session and, therefore, a total of 480 trials at the end of the two sessions. Each face was opposed to a random vehicle. In one session, the target stimulus was the face (vehicle distractor), while in the other session, the target stimulus was the vehicle (face distractor). A calibration phase was performed at the beginning and middle of each session and a drift correction was applied every 10 trials (if the drift was larger than 1°, then recalibration was performed). During the calibration phase, participants were asked to gaze at nine white dots appearing sequentially in a 3 × 3 grid covering the entire screen. Matlab (MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997) were used to control timing and stimulus display as well as communication with the eye-tracker.

During the experiment, participants were seated on an adjustable chair in a semi-lighted room. The head was stabilized by means of a forehead and a chin-rest at a fixed distance of 57 cm from the screen. A session lasted approximately 20 min and the whole experimental procedure took approximately 50 min. The target category (face or vehicle) was defined before each session. At the beginning of each trial, participants were asked to fixate a white cross during a pseudo-random time interval ranging between 800 and 1600 ms. After a 200-ms gap, two images (a face and a vehicle) were simultaneously displayed on each side of the screen for 400 ms (Fig. 2), and participants were asked to make a saccade as fast as possible toward the target image. The center of each image was located at a fixed distance of 8° of eccentricity from the center of the screen. Each trial ended with the presentation of a gray background for 1000 ms.

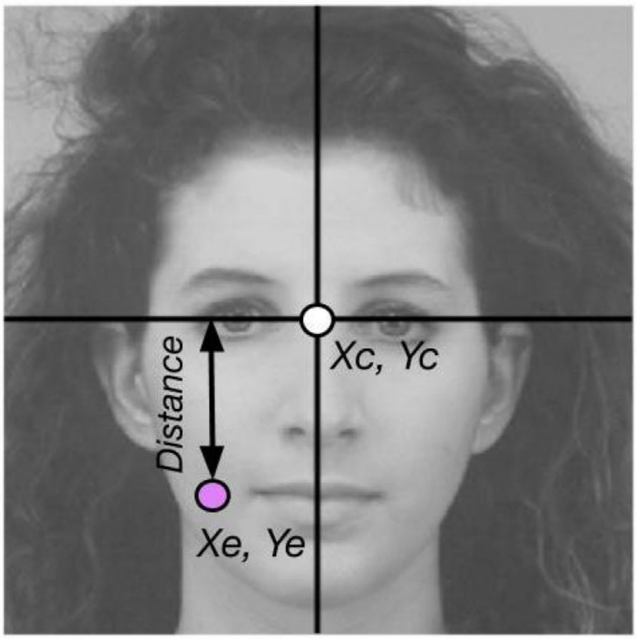

Fig 3. Visual representation of the vertical distance to the center. X_c,Y_c denote the central point of the image, and X_e,Y_e denote the endpoint of the saccade. The vertical distance to the center corresponds to the difference between Y_c and Y_e.

2.1.4. Data analysis

Before any further analysis, the eye movement data were preprocessed in order to eliminate trials that we considered to be invalid. Valid trials were selected according to the following validity criteria. First, a saccade had to be the first event after stimulus onset, with no blink occurring during its execution. Second, this first saccade had to have a latency greater than 50 ms (to avoid anticipatory saccades), a starting point within a radius of 2° around the center of the screen, and a duration smaller than 100 ms. Moreover, the saccade amplitude had to be greater than 1° and should not go beyond the screen. This preprocessing led to 10.4% of the initial number of trials being rejected. For all valid trials, only the first saccade and fixation were analyzed.

Statistical analyses were carried out using the open-source software R (R Core Team, 2016) with R Studio 1.1.456 (Racine, 2012). A saccade was considered as "correct" if it was directed toward the side of the display containing the target and as an "error" if it was in the opposite direction (i.e., directed toward the distractor).

In order to quantify the orienting of attention toward local face features, such as the mouth or eyes, we computed saccade endpoints. To do this, we extracted and visualized the coordinates X_e and Y_e of the first saccade endpoint in the image space (i.e., within a square of 11 × 11 degrees, the coordinates X_0 and Y_0 being at the top-left corner). Overall, saccades tended to land around the eyes, which corresponded to the center of the image. As a measure to compare endpoint positions between conditions, we analyzed the vertical distance between the endpoint of the first correct saccade and the center of the image in each condition. This distance corresponds to the distance between the Y-coordinate of the endpoint Y_e of the saccade

and the Y-coordinate of the center of the image Y_c. A visual representation of the distance to the center is presented in Fig. 3, and a negative value corresponds to a saccade landing below the image center. This measure was also computed for the vehicles.

Mean accuracy (in % of correct responses), mean first saccade latency (in ms), and mean vertical distance to the image center (in degrees of visual angle) were computed for each participant in each experimental condition and analyzed as dependent variables. First, a paired samples *t*-test with the Target (Face, Vehicle) as a within-subject factor was used to assess the main effect of the target. Next, a repeated measures ANOVA with the Emotional Facial target (EFE; Happy, Neutral, or Fearful) as a within-subject factor was conducted for saccades when the target was a face. Similarly, a repeated measures ANOVA with the Emotional Facial distractor (EFE; Happy, Neutral, or Fearful) as a within-subject factor was also conducted for saccades when the target was a vehicle. If needed (i.e., if a significant effect of the EFE target or EFE distractor was observed), paired samples *t*-tests were used for pairwise comparisons between Emotional Facial Expressions (EFE; Happy, Neutral, or Fearful). Effect sizes were estimated by calculating partial eta-squared ($\eta_p{}^2$) for ANOVAs and Cohen's d for *t*-tests. An effect was considered significant if its *p* value was below the threshold $\alpha = .05$.

Before performing the parametric tests, statistical assumptions were tested using a K-S corrected Lilliefors test (Lilliefors, 1967) for normality of distributions (of within-pair differences for *t*-tests and of the variable for repeated measures ANOVAs; McCrum-Gardner, 2008), and Mauchly's sphericity test was used to test for equality of variances (of the differences between all possible pairs; for repeated measures ANOVAs). When distributions deviated significantly from the normal distribution ($p < .05$), nonparametric tests were used. More precisely, Wilcoxon Signed-ranks tests and Friedman tests were performed on the dependent variables instead of the *t*-tests and ANOVAs, respectively. Therefore, when the conditions permitted the use of parametric tests, we favored the use of such tests, as they are known to be more powerful than nonparametric tests (Hoskin, 2012).

Finally, we computed the minimum latency (also referred to as the minimum saccadic reaction times in previous studies—Crouzet et al., 2010; Guyader et al., 2017; Kauffmann et al., 2021—for each Target condition and each facial target EFE). To compute the minimum latency, we recorded the latencies of all the first saccades for all participants. We computed their distribution while taking account of the saccade accuracy (Correct, Error), the type of target (Face, Vehicle) and, in the case of the face targets, the type of EFE (Neutral, Happy, or Fearful). For face and vehicle targets, 13,194 (1863 errors) and 13,037 (2125 errors) saccades, respectively, were used to compute the distributions. For neutral, happy, and fearful faces, the distributions contained 4394 (640 errors), 4416 (601 errors), and 4384 (622 errors) saccades, respectively. The minimum latency corresponds to the time as of which there were significantly more correct than error saccades. More precisely, distributions were divided into 10-ms time bins (e.g., the 170-ms bin contained latencies from 165 to 174 ms), and for each bin, we used a $\chi 2$ test (with a criterion of $p < .05$) to test if there were significantly more correct than error saccades. If there were significantly more correct than error saccades in five consecutive bins, the first of these bins was defined as the minimum latency. Note that this procedure was the same as in previous papers using saccadic choice tasks (e.g., Crouzet et al.,

2010; Guyader et al., 2017; Kauffmann et al., 2021). Experimental data, analysis code, and simulation code are available at https://osf.io/bjmcy/.

2.2. Results

2.2.1. Accuracy

A paired samples *t*-test performed on mean accuracy (Fig. 4a) indicated a significant effect of the Target ($t(60) = 2.52$, $p = .014$, $d = 0.32$). Participants were more accurate when the target was a face ($M \pm SD$: .88 ± .076) than when it was a vehicle ($M \pm SD$: .86 ± .082). Neither an effect of the EFE for facial targets nor an effect of the EFE distractor for vehicles targets was found.

2.2.2. Latency

A paired samples *t*-test performed on mean saccade latency (Fig. 4b) indicated a significant effect of the Target ($t(60) = –5.43$, $p < .001$, $d = 0.69$). Saccades were elicited faster when the target was a face ($M \pm SD$: 176 ± 21.5 ms) than when it was a vehicle ($M \pm SD$: 191 ± 27.2 ms). Neither an effect of the EFE for facial targets nor an effect of the EFE distractor for vehicles targets was found.

2.2.3. Minimum latency

The minimum latency (Fig. 4d) was found in the 110-ms bin for faces (overall, and also for neutral, happy, and fearful faces independently) and in the 130-ms bin for vehicles.

2.2.4. Endpoints

A Wilcoxon Signed-ranks test performed on mean vertical distance to the image center (Fig. 4c) revealed a main effect of the Target ($Z = 5.6$, $p < .001$). Saccades landed higher when the Target was a vehicle ($Mdn = –0.016°$) than a face ($Mdn = –0.32\ °$). A nonparametric Friedman test revealed a main effect of the EFE ($F(2) = 50.6$, $p < .001$) when the target was a face. In this condition, saccades landed higher for neutral faces ($M \pm SD$: –0.26 ± 0.61°) than for fearful ($M \pm SD$: –0.29 ± 0.61°; $t(60) = 3.78$, $p < .001$, $d = 0.92$) or happy ($M \pm SD$: –0.38 ± 0.61 °; $t(60) = –9.04$, $p < .001$, $d = 1.15$) faces. Moreover, saccades toward fearful faces also landed higher than those toward happy faces ($t(60) = –7.21$, $p < .001$, $d = 0.48$). Finally, we did not find any effect of the EFE distractor when the target was a vehicle. Fig. 5 presents examples of heat maps computed based on the saccade endpoints of all subjects. For this representation, we chose to display the heat maps on the top of three randomly chosen faces: a neutral, a happy, and a fearful face. Heat maps were obtained by (1) adding all the first correct saccades and (2) convolving a small 2D Gaussian on each endpoint.

2.2.5. Bayes factor analysis to test the lack of emotional modulation on very fast saccades

Results of Experiment 1 did not show any significant effect of the EFE (Happy, Fearful, and Neutral) on saccade accuracy and saccade latency when the target was a face. Therefore, the null hypothesis (H_0: no effect of the EFE on mean saccade accuracy or mean saccade latency) cannot be rejected and no conclusion can be drawn (Hoijtink, Mulder, van Lissa,

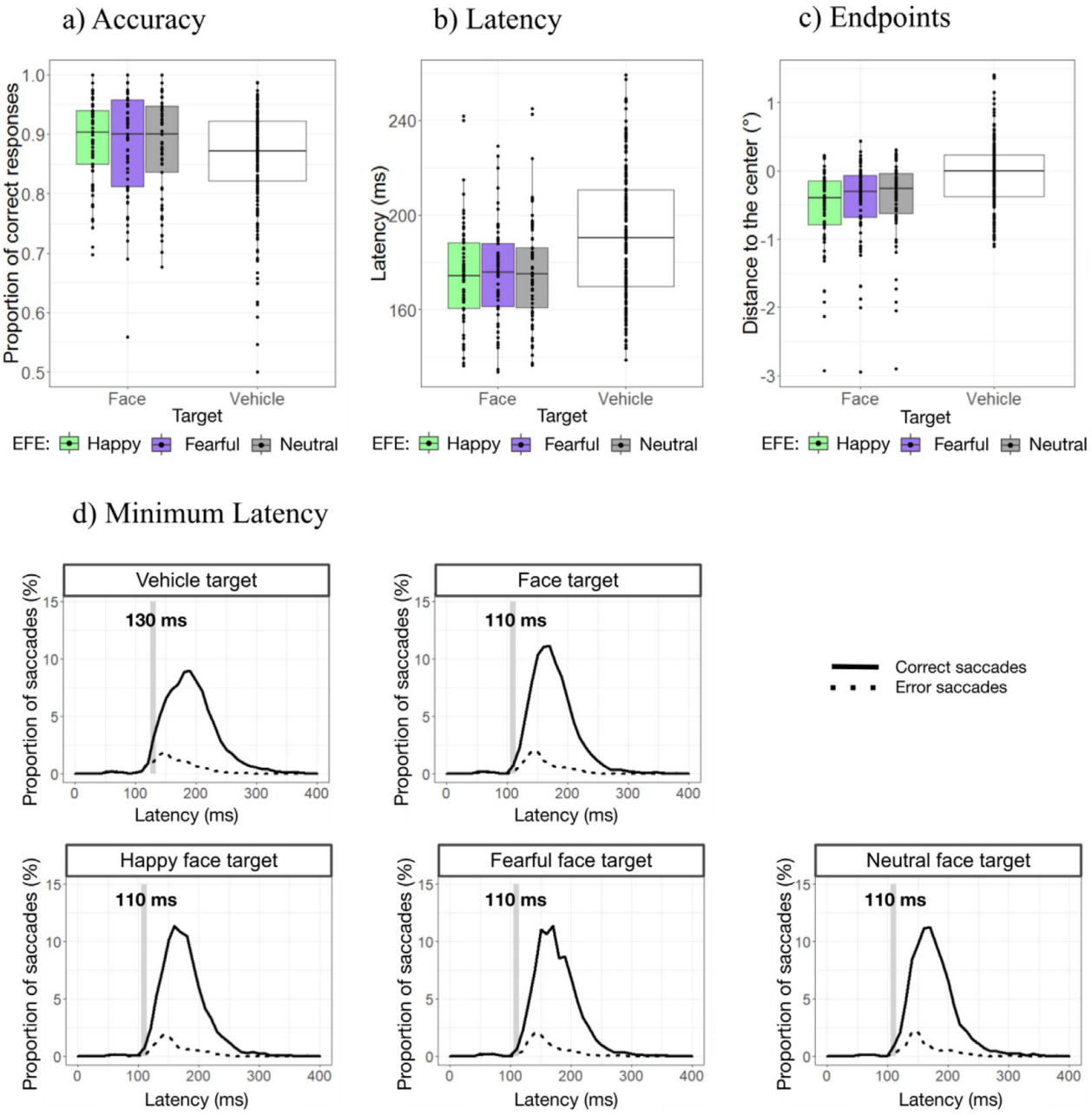

Fig 4. Boxplots for (a) mean proportion of correct responses, (b) mean latency (in ms), and (c) mean distance to the image center (in degrees of visual angle), according to the Target (Face and Vehicle) and the Emotional Facial Expression of face targets (EFE; Happy, Fearful, or Neutral). (d) Distribution of saccade latencies for each Target and each Emotional Facial Expression of face targets. Unbroken lines correspond to correct saccades and dotted lines to error saccades. The gray bar corresponds to the 10-ms bin containing the minimum latency. It should be noted that for the purposes of illustration, and because no significant effect of the EFE distractor was found, all types of face distractors (happy, fearful, and neutral) were recorded for vehicle targets.

& Gu, 2019; Wagenmakers, 2007). To evaluate the probability of the presence or absence of an effect of EFE, we used a method based on Bayesian statistics (Bayes factors; Kass & Raftery, 1995). This method was added to the previous analyses to evaluate the probability of H_0 more precisely compared to the alternative hypothesis (H_a: not H_0). This was done for

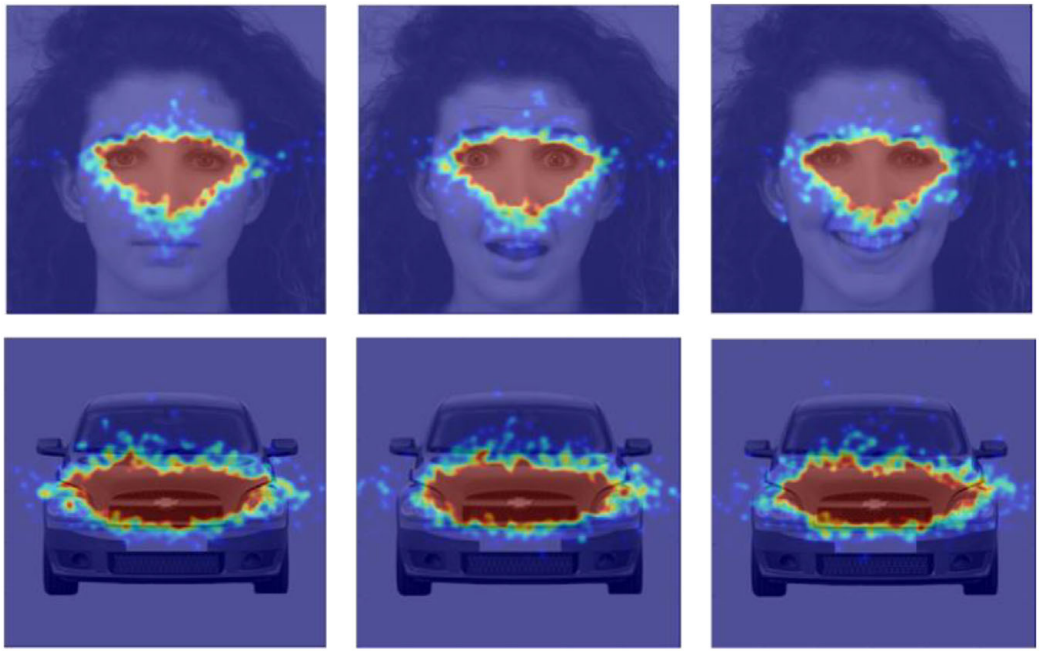

Fig 5. Heat maps computed from the endpoints of all the first correct saccades toward neutral (top left), fearful (top center), or happy (top right) faces, and toward vehicles when the distractor was a neutral (bottom left), fearful (bottom center), or happy (bottom right) face.

saccade accuracy and latency independently. We used the bain (Bayesian informative hypotheses evaluation; Gu, Hoijtink, Mulder, & Rosseel, 2019; Hoijtink et al., 2019) R package. With this package, the variance of the prior distribution for each of the means is computed using a fraction of the information in the data for each group mean (which here renders a prior variance of 0.01 and 706 for the accuracy and latency, respectively).

An ANOVA was computed to estimate the mean accuracy and mean latency when the target was a face in each of the three emotion conditions, happy (*accuracy*, $M \pm SD$: 0.89 ± 0.072; *latency,* $M \pm SD$: 175 ± 21.6 ms), fearful (*accuracy*, $M \pm SD$: 0.88 ± 0.091, *latency,* $M \pm SD$: 175 ± 20.6 ms), and neutral (*accuracy*, $M \pm SD$: 0.88 ± 0.084, *latency,* $M \pm SD$: 176 ± 22.7 ms). Two hypotheses were evaluated:

$$H_0 : M_{Happy} = M_{Fearful} = M_{Neutral}$$

$$H_a : \text{not } H_0$$

where M_{Happy}, $M_{Fearful}$, and $M_{Neutral}$ denote the mean accuracy or mean latency for happy, fearful, or neutral face targets.

For accuracy, the Bayes factor versus H_a was 74, and the posterior probabilities (computed assuming equal prior probabilities) were 0.99 for H_0 and 0.01 for H_a. This Bayes factor suggests that the data are 74 times more likely to occur under H_0 than under H_a, and can be interpreted as providing very strong support for H_0 (for a scale for interpretation of the Bayes factor, see Jeffreys, 1998).

For latency, the Bayes factor versus H_a was 88.3, and the posterior probabilities (computed assuming equal prior probabilities) were 0.99 for H_0 and 0.01 for the $H_{a.}$. This Bayes factor suggests that the data are 88 times more likely to occur under H_0 than under H_a, and can also be interpreted as providing very strong support for H_0 (Jeffreys, 1998).

2.3. Discussion

Results of Experiment 1 replicate previous findings showing that participants made fewer errors and initiated saccades faster when the target was a face than when it was a vehicle. Accuracy for face targets in our experiment was similar to that reported in previous saccadic choice tasks with face and vehicle targets (88% correct responses on average in this study, compared to 89.6%, 86%, and 87.5% for previous studies using a saccadic choice task with face-vehicle pairs, respectively, Crouzet et al., 2010; Guyader et al., 2017; Kauffmann et al., 2019). However, accuracy for vehicle targets in this study was higher (86% correct responses on average in this study, compared to 71%, 71%, and 76.6% for previous studies using a saccadic choice task with face-vehicle pairs, respectively, Crouzet et al., 2010; Guyader et al., 2017; Kauffmann et al., 2019). The same pattern (similar results for face detection and better detection of vehicle targets in this experiment compared to previous ones) was found for the mean and minimum latencies. Furthermore, a previous study (Crouzet et al., 2010) had observed a tendency of early saccades (100–140 ms) to go toward the side with the faces, even if the task required a saccade toward the vehicles. However, no such effect was found here.

The better detection of vehicle stimuli in our experiment compared to previous ones may be due to the fact that we used more prototypical images with no background, thus reducing the variability between stimuli. Moreover, vehicle stimuli were duplicated to correspond to the three emotions of the same face (the same vehicle was presented three times), thus contributing to their low variability. It can also be noted that the high degree of within-category homogeneity of faces and cars resulted in both categories having distinct amplitude spectrum (AS) properties. Previous studies have shown that such information is used during the saccadic choice task and could partly explain the bias for faces (Crouzet & Thorpe, 2011; Honey et al., 2008). For example, Honey et al. showed that images of faces for which the phase of the Fourier component (i.e., spatial relations within the image) was disrupted, while the AS was preserved still elicited faster saccades than images of vehicles with similar alterations. A recent study (Kauffmann et al., 2021), however, found that faster saccades toward faces than vehicles could be observed even when the AS of the stimuli was made more similar for faces and cars, suggesting that AS differences between faces and cars cannot entirely explain the bias in favor of faces.

Concerning facial expressions, no effect was observed either on mean latency or on accuracy. The Bayes factor analysis provides an interesting statistical tool allowing us to draw inferences about the likelihood of a null effect of emotions on saccades toward faces. The conclusion of Experiment 1 is that there is a very strong evidence that emotions do not influence fast saccades toward faces in a saccadic choice task in which participants have to saccade toward a face when a face and a vehicle are simultaneously displayed. Nevertheless, an effect

of facial expressions was observed on saccade endpoints. Indeed, overall, saccades tended to land around the eyes, but landed higher when the face was neutral than when it was fearful or happy and also when the face was fearful rather than happy. The fact we found no prioritization of emotional faces contrasts with previous studies suggesting that emotional, and especially threatening, events are automatically (i.e., rapidly and nonintentionally) prioritized (Öhman, 2005). Fast face detection could be the result of "quick and dirty processing" that may be insufficient to decode expressions (Crouzet & Thorpe, 2011). Therefore, one explanation may be that expressions are not yet decoded at such small latencies and before a face is detected (Kulke, 2019; Mulckhuyse, 2018; Schyns et al., 2009).

In the second experiment, we used the same experimental design but with pairs of faces, one emotional (happy or fearful) and the other neutral, in order to directly test the detection of facial expressions. In one session, participants had to saccade toward emotional faces, and in the other, they had to saccade toward neutral faces. We expected to find better performances when the target was an emotional than when it was a neutral face. Furthermore, assuming that humans evolved to preferentially orient their attention toward threat, we should observe better performances toward fearful than happy face targets. However, behavioral studies showing an advantage for threatening faces have used manual responses and may reflect processes occurring at a later stage of visual processing. Thus, even though fearful faces were not prioritized in Experiment 1, we can still suggest that attention may be preferentially attracted by threat if we consider that this prioritization occurs at a later stage. Alternatively, assuming that happy faces are easier to detect, we should find better performances for happy faces. We still expected saccade endpoints to be higher for fearful than for happy faces.

3. Experiment 2: Emotional face versus neutral face

3.1. Materials and methods

3.1.1. Participants

Twenty participants (nine females; mean age ± *SD*: 23.95 ± 5.26 years; age range: 19–41 years) were recruited from the local university to perform a saccadic choice task. All participants had normal or corrected-to-normal visual acuity. Undergraduate psychology students received course credits for their participation in the experiment. All participants gave their informed written consent before the experiment, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

3.1.2. Stimuli

Stimuli were 180 grayscale photographs of emotional and neutral faces chosen from the KDEF database (Fig. 6). More precisely, photographs of 60 different individuals (30 females) displaying three different facial expressions (neutral, happy, and fearful expressions) were included. All individuals presented in Experiment 1 were included in Experiment 2. Also, neutral stimuli were duplicated twice in this experiment to counterbalance the presentation of

Fig 6. Example of emotional and neutral face stimuli used in Experiment 2 (left) and their mean amplitude spectrum (right). Each emotional face (top) is associated with a neutral face (bottom).

neutral and emotional stimuli. This led to a total of 240 images. The images again sized 300 × 300 pixels, corresponding to 11 × 11 degrees of visual angle, and were equalized in terms of mean luminance and root mean squared contrast (mean values of 126 and 66, respectively, for pixel intensity values between [0, 255], based on mean luminance and contrast values of all the stimuli) before a gamma correction was applied. A training session with eight practice trials involving four additional individuals was performed at the beginning of the experiment to allow participants to familiarize themselves with the task.

3.1.3. Procedure

The procedure was the same as in the first experiment. In each session, a neutral face was always displayed together with an emotional (happy or fearful) face, and participants were asked to make a saccade as fast as possible toward the target (the emotional or the neutral face), which was defined at the beginning of each session. A trial began with the presentation of a central fixation cross for a pseudo-random time period ranging between 800 and 1600 ms. After a 200 ms gap, an emotional and a neutral face were randomly and simultaneously displayed on either side of the screen for 800 ms (Fig. 7). In each session, each of the 240 different images was displayed, once on the left and once on the right side of the screen, leading to a total of 240 trials per session. The trial ended with the presentation of a gray screen for 1000 ms. One male face and one female face were displayed for each trial.

3.1.4. Data analysis

Preprocessing and data analysis methods were the same as in Experiment 1. The preprocessing procedure led to 7.8% of the initial number of trials being rejected. Mean accuracy, mean first saccade latency, as well as mean distance to the center of the image (which again corresponded to the eyes) were computed for each participant in each experimental condition

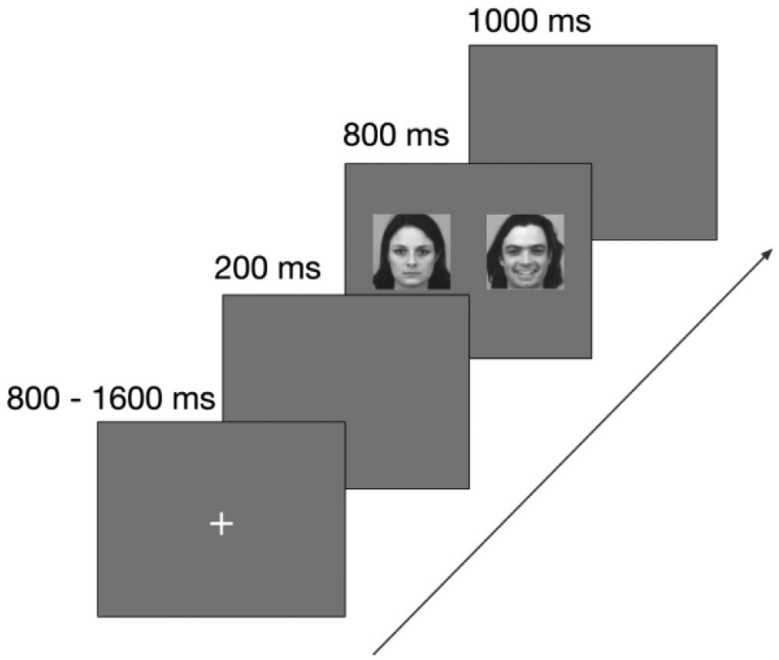

Fig 7. Time course of a trial for Experiment 2.

and analyzed as dependent variables. First, a paired samples *t*-test with the Target (Emotional or Neutral) as a within-subject factor was used to assess the main effect of the Target. Then, a paired samples *t*-test with the Emotional Facial target (EFE; Happy or Fearful) as a within-subject factor was applied for saccades in cases when the target was an emotional face. A paired samples *t*-test with the Emotional Facial distractor (EFE; Happy or Fearful) as a within-subject factor was applied for saccades in cases when the target was a neutral face.

Again, statistical assumptions were tested using a K-S corrected Lilliefors test for normal distribution of within-pair differences. When distributions deviated significantly from the normal distribution ($p < .05$), a Wilcoxon Signed-ranks test was applied. Minimum latency was computed in the same way as in the first experiment. For emotional and neutral targets, 4387 (1390 errors) and 4450 (1664 errors) saccades, respectively, were used to compute the distributions. For happy and fearful faces, the distributions contained 2196 (647 errors) and 2191 (743 errors) saccades.

3.2. Results

3.2.1. Accuracy

Paired samples *t*-tests performed on mean accuracy (Fig. 8a) indicated a significant effect of the Target ($t(19) = 3.91$, $p < .001$, $d = 0.87$) and a significant effect of the EFE target when the Target was an emotional face ($t(19) = –3.55$, $p = .002$, $d = 0.79$). Participants made more correct saccades (i.e., first saccades toward the target) when they were asked to saccade toward the emotional ($M \pm SD$: 0.68 ± 0.11) rather than the neutral face ($M \pm SD$: $.62 \pm .13$), and also when the emotional Target face was happy ($M \pm SD$: $.70 \pm .11$) rather

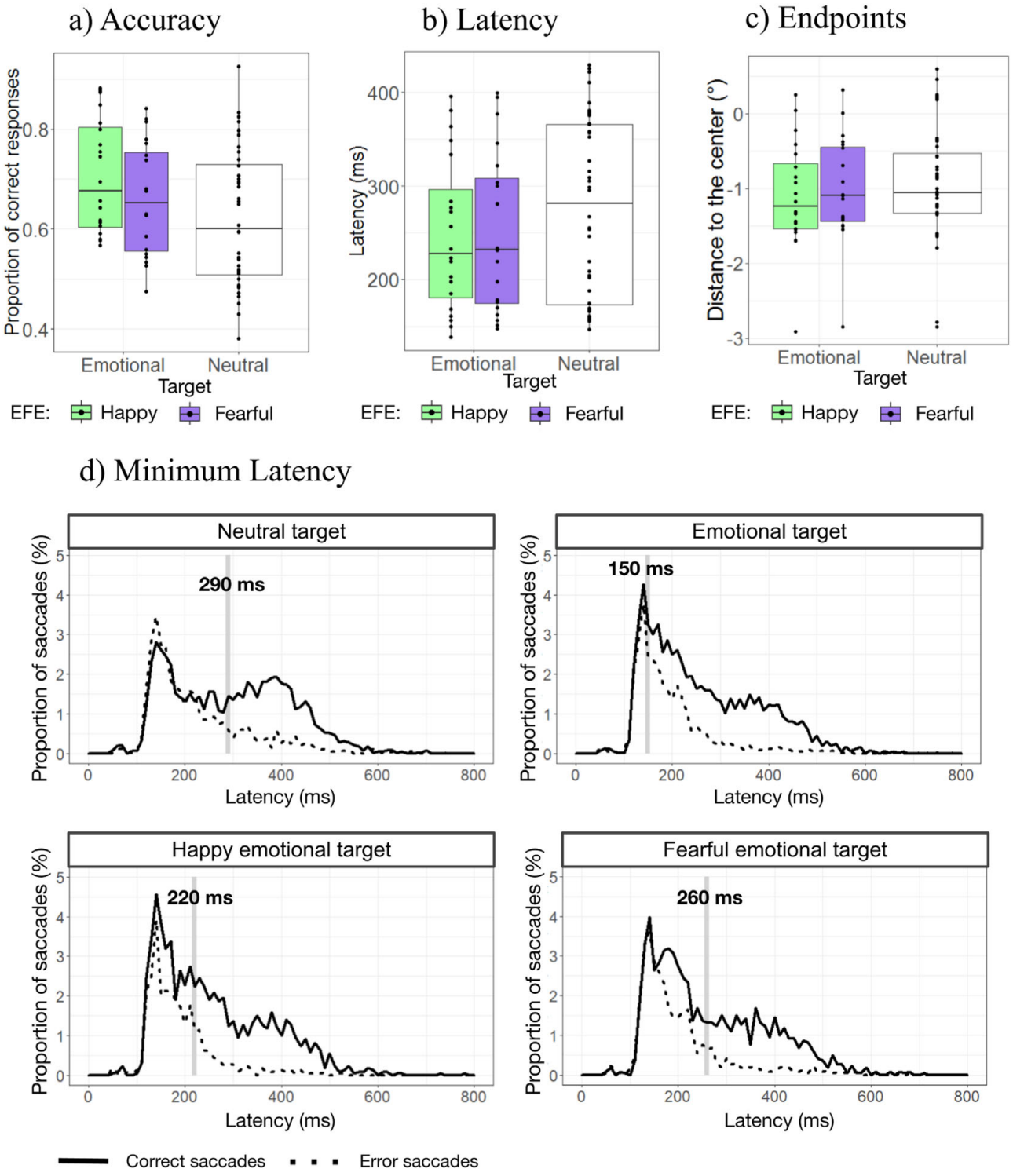

Fig 8. Boxplots for (a) mean proportion of correct responses, (b) mean latency (in ms), and (c) mean distance to the image center (in degrees of visual angle) for correct saccades as a function of the Target (Emotional and Neutral) and the Emotional Facial Expression of emotional targets (EFE; Happy or Fearful). (d) Distribution of saccade latencies for each Target, and for each Emotional Facial Expression of the emotional targets. Unbroken lines correspond to correct saccades and dotted lines to error saccades. The gray bar corresponds to the 10-ms bin containing the minimum latency. It should be noted that, for the purposes of illustration and because no significant effect of the EFE distractor was found, all types of emotional face distractors (happy or fearful) were recorded for neutral targets.

than fearful ($M \pm SD$: .66 ± .11). Finally, we did not find any effect of the EFE distractor when the Target was neutral.

3.2.2. Latency

Paired samples *t*-tests performed on mean latency (Fig. 8b) showed only a significant effect of the Target ($t(19) = -3.16$, $p = .005$, $d = 0.71$). Saccades were elicited faster when participants were asked to saccade toward an emotional face ($M \pm SD$: 249 ± 80.1 ms) rather than a neutral face ($M \pm SD$: 277 ± 96.1 ms). No effect of the EFE was found for emotional targets and there was also no effect of the EFE distractor for neutral targets.

3.2.3. Minimum latency

The minimum latency (Fig. 8d) was located in the 150-ms bin for emotional faces and in the 290-ms bin for neutral faces. For happy emotional face targets, the minimum latency was 220 ms, and for fearful emotional face targets, it was 260 ms.

3.2.4. Endpoints

Paired samples *t*-tests performed on the mean vertical distance between the image center and the endpoint (Fig. 8c) revealed only a main effect of the EFE for emotional face targets ($t(19) = -3.68$, $p = .002$, $d = 0.82$). Saccades landed higher when the emotional face was fearful ($M \pm SD$: –1.02 ± 0.7°) than when it was happy ($M \pm SD$: –1.11 ± 0.72°). We did not find any effect of the EFE distractor for neutral targets. Fig. 9 presents examples of computed heat maps displayed on top of a randomly chosen face for each emotional condition.

3.3. Discussion

Results of Experiment 2 confirm that emotional faces are easier to detect than neutral faces during a saccadic task in which two faces are simultaneously displayed (Bannerman, Milders, & Sahraie, 2009). First, saccades were more often correct when the target was the emotional face. This was especially the case when the emotional face was happy rather than fearful. Second, saccades were also elicited faster when the target was the emotional face. Surprisingly, however, there was no significant effect of facial expressions on mean saccade latencies, despite the fact that the minimum latency was higher for fearful than happy faces. The distribution of saccade latencies for correct and error saccades in Experiment 2 was quite different from that found in Experiment 1. First, there were more errors in Experiment 2, and we can see that most of the error saccades had short latencies (e.g., below 200 ms) and that they decreased in number as latencies increased. Reliable saccades occurred at 150 ms for emotional targets and 290 ms for neutral targets. Moreover, if we consider the distribution of responses in the 140–160 ms time windows, saccades tended to be made toward the emotional faces even when the target was the neutral face. It is, therefore, possible that saccades toward emotional faces are harder to control in this time window. Also, participants might have adopted the strategy of detecting the emotional face first and then deducing the position of the neutral face from this. Since the open mouth is more salient, it might be easier to detect and this would justify such a strategy. With regard to the saccade endpoints, these were still

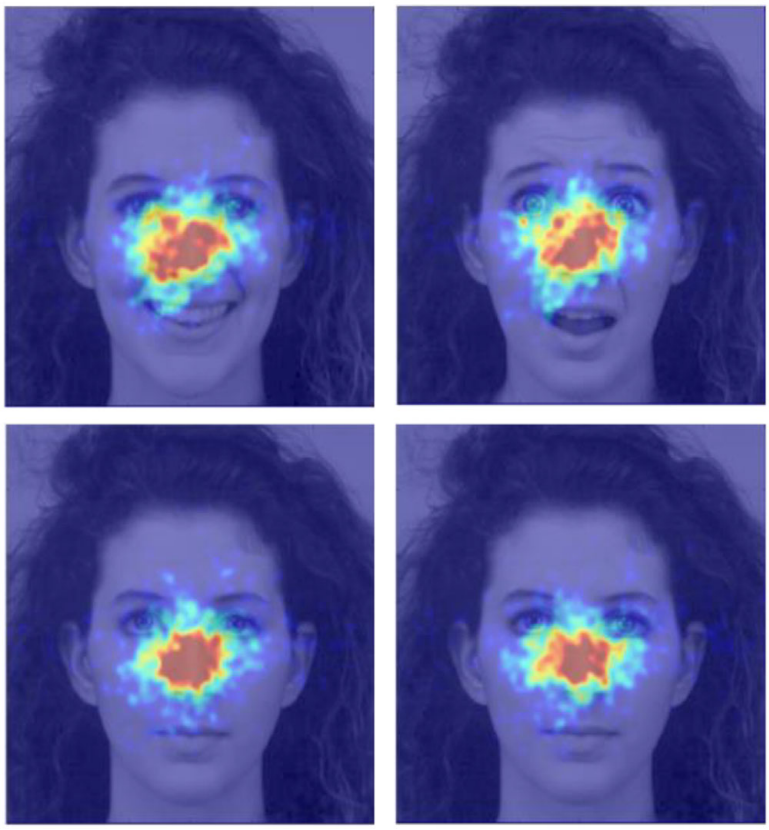

Fig 9. Heat maps computed from all first correct saccade endpoints in the second experiment when the target was a happy face (top left), a fearful face (top right), a neutral face when the distractor was happy (bottom left), or a neutral face when the distractor was fearful (bottom right).

higher when the target was a fearful rather than a happy face. Compared to the first experiment (in which saccades tended to land around the eyes), saccades tended to land around the nose in Experiment 2. This might suggest that even if local features are able to capture attention in an automatic way, the weight allocated to those features can also be modulated by the task.

Overall, it is likely that a parsimonious hypothesis related to the simple perceptual properties of the stimuli is capable of explaining the happy face advantage that we observed in this experiment, without reference to emotional processes (Calvo & Nummenmaa, 2011). Fearful and neutral faces may be statistically more similar than happy and neutral faces. Using computational models, it has already been shown that in categorization tasks, happy faces are easier to recognize (i.e., elicit a higher rate of correct recognition) than fearful faces, and that happy faces are more different from neutral than from angry faces (Dailey, Cottrell, Padgett, & Adolphs, 2002; Mermillod et al., 2009). In the next section of the article, we describe an artificial neural network used to quantify the perceptual differences between neutral and fearful or happy faces. Even if we did not directly simulate saccadic responses, the task was nevertheless similar, as it involved finding the location of an emotional compared to a neutral face. More precisely, the neural network was trained and tested on its ability to discriminate between emotional-neutral and neutral-emotional face pairs (with the emotional face being

either on the right or left side of the pair). Therefore, this task can be considered as a categorization task, as it makes it necessary to classify face pairs on the basis of two categories: the neutral-emotional category and the emotional-neutral category. Hence, its goal was to decide whether the emotional face was on the left or right side of the pair. Next, the network was tested and we computed its performance in the correct categorization of pairs of faces. The results were then further subdivided depending on whether the emotional face on the pair was happy or fearful. The network's performance might have been found to be better when the emotional face was fearful or when it was happy or it might have been the same in the two conditions. Based on results from Experiment 2, our hypothesis was that the network would be able to discriminate the emotional and neutral faces better when the emotional face was happy.

4. Simulations

The whole simulation procedure was very similar to previous studies that have used a multilayer perceptron (MLP) for emotion categorization (Dailey et al., 2002; Mermillod et al., 2009, 2010, 2019). It can, therefore, be subdivided into two steps: a preprocessing step in which Gabor filters are applied to the Fourier transform of the overall image, and a training-testing procedure using the MLP. We opted for this design (e.g., instead of a convolutional neural network) because previous studies have shown that, even though the method is less efficient for artificial intelligence purposes, the use of Gabor filters tuned at different spatial frequencies and orientation channels permits a more biologically plausible simulation of the primary visual cortex (Jones & Palmer, 1987) as well as of the phase invariance properties related to V1 complex cells (Hubel & Wiesel, 1968). Moreover, it produces results that are very similar to humans for the same facial emotion recognition task (Dailey et al., 2002; Li & Cottrell, 2012).

4.1. Method

4.1.1. Preprocessing and stimuli

There are many ways to reduce the size of images, but as we wanted to compare the performance of the network with that of our participants, we chose to do this in a biologically plausible way. Indeed, images were described using a bank of Gabor filters applied in the frequency domain. Each filter simulated the functioning of primary visual cells by being sensitive to a particular spatial frequency band and a particular orientation (Hubel & Wiesel, 1968; Jones & Palmer, 1987). Each image was described in terms of its energy at each filter output. More precisely, preprocessing began with the application of a Hanning window on each image to avoid boundary effects. The images were then transferred to the Fourier domain and Gabor filters were applied on the overall image. Therefore, each filter indicated the amount of energy in the image for one frequency channel and one orientation. We used a bank of 48 Gabor filters tuned to six spatial frequency channels (central frequencies = {0.82; 1.23; 1.85; 2.78; 4.14; 6.25}, given in cycles/degree) and eight orientations (0, pi/8, 2pi/8,

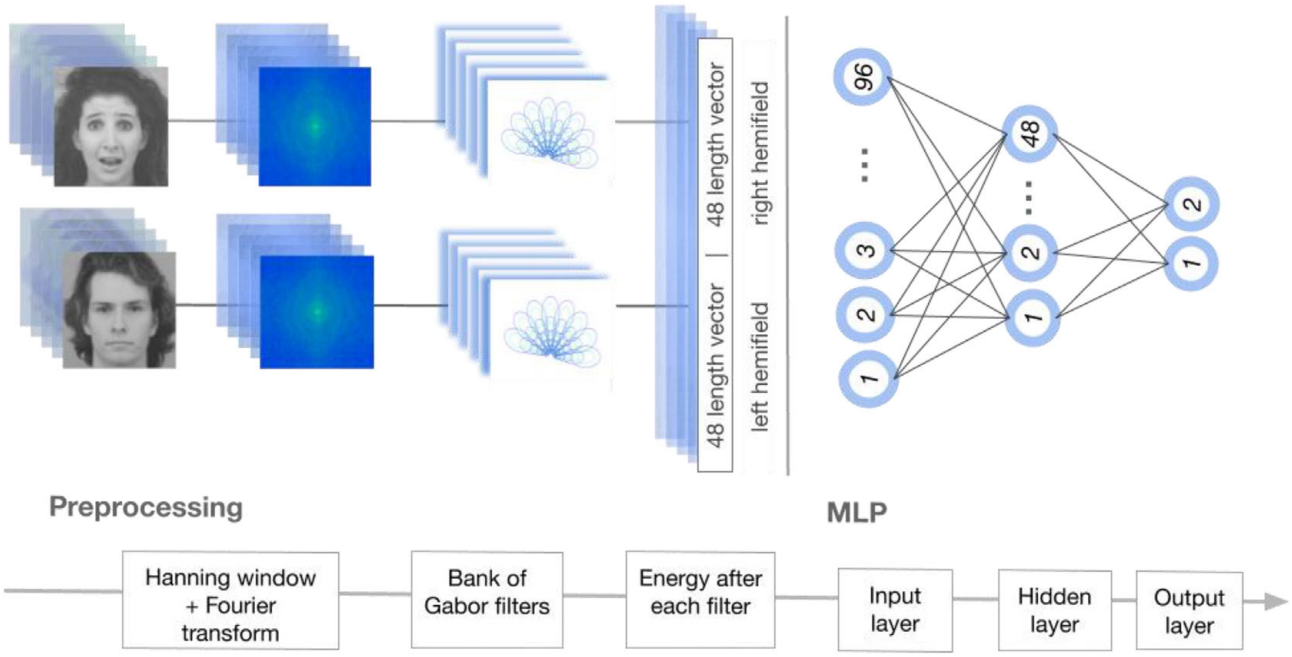

Fig 10. Representation of preprocessing steps and neural network (three-layer MLP with fully interconnected neurons). The module for the Fourier transform is presented for the purposes of illustration.

3pi/8, 4pi/8, 5pi/8, 6pi/8, 7pi/8). The 48 filters were applied to each image and the energy of each filtered image was computed, resulting in an energy vector with 48 values. Each value corresponded to the local energy spectra of the image in the spectral domain multiplied by the kernel of the Gabor filter in a specific spatial frequency band at a specific orientation. Finally, each image was described by a 48-length vector. These values were normalized between 0 and 1 across all faces and emotions. Inputs were fed into an MLP, whose task was to associate the descriptor of the face with the output vector of the category.

In order to match the experimental procedure of Experiment 2, we gathered the image vectors together in pairs corresponding to all the possible combinations that could occur in Experiment 2. Consequently, these pairs were necessarily composed of faces of different genders and with different facial expressions: neutral on one side of the vector and emotional (either happy or fearful) on the other side. The association of the vectors was computed by simply concatenating the two vectors, and each pair was represented twice, one with the emotional face on the right and the other with the emotional face on the left. Each pair was also associated with a specific label, which can be considered to represent the probability of each category (the emotional-neutral and the neutral-emotional category): [1, 0] if the emotional face was on the left and [0, 1] if the emotional face was on the right. A visual representation of the preprocessing steps and neural network is presented in Fig. 10.

4.1.2. Network architecture

The purpose of the network was to simulate a discrimination task, similar to the one presented in Experiment 2, involving a simultaneously presented emotional (happy, fearful) and neutral face. Therefore, the input was not simply a single vector corresponding to a single

image, but a combination of two vectors corresponding to a pair of images. The network was trained to discriminate emotional-neutral face pairs from neutral-emotional face pairs and its architecture consisted of three different layers with 96 input units, 48 hidden units, and 2 output units. On the last layer, a standard sigmoid transfer function was used as an output function, given by:

$$f\ (x) = \frac{1}{1 + e^{-x}}\ ,$$

where x is the weighted input to the layer (i.e., the sum of the layer input multiplied by the weight matrix, which has random initial weights). The standard backpropagation algorithm was used for synaptic weight adjustment during training, with a learning rate set to 0.01, and the Adam algorithm (Kingma & Ba, 2017) for optimization. The error signal used for synaptic weight correction was computed based on the mean squared error (MSE):

$$MSE\ = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

where n is the number of values in the vector, y is the expected output, and $\hat{y}$ is the actual output.

4.1.3. Procedure

At the beginning of the procedure, 40 of the 60 different faces from Experiment 2 (20 women) were randomly selected for training and 20 of them (10 women) for testing. Thus, none of the tested faces were used in the training phase and they were, therefore, unknown to the network. We picked the corresponding vector pairs among all possible vector pairs. As we did not want any individuals to be present in both the training and testing phase, no pairs that contained both a training and a test individual were used. This led to the selection of 800 pairs for testing and 3200 pairs for training for each training-test procedure.

After selecting the training-test sets, the 96-length energy vectors corresponding to the training set were fed into the network using the standard backpropagation algorithm. The network associated the training input vectors with the corresponding output vector over 500 iterations. All pairs were forwarded and backpropagated to the network at the same time and the gradient was thus computed using the whole training dataset (i.e., a batch gradient descent method was used). After learning, the network was tested on 800 new pairs of faces. The output from the model was a vector with two values, corresponding to the probabilities of the pair being associated with the emotional-neutral category and the neutral-emotional category. The network classification response was then assigned to the class with the highest probability (i.e., a winner-takes-all procedure was applied), and the accuracy was set to 1 if the expected and actual responses were the same, and 0 if not. The same training-test procedure was repeated over 50 iterations in order to calculate a stable and reliable average accuracy from 50 different networks.

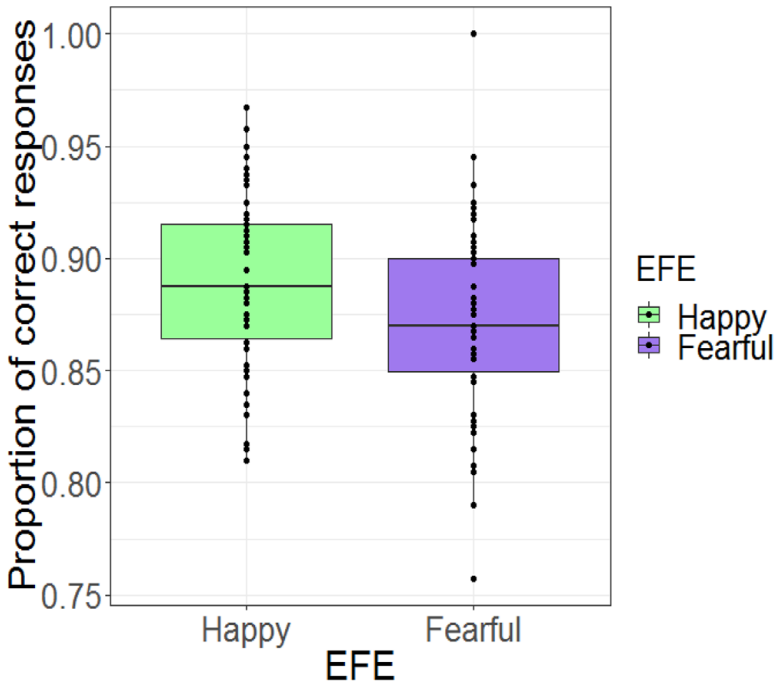

Fig 11. Boxplot for mean accuracy of the model depending on the EFE of the emotional face (happy, fearful).

4.2. Results

On average, the network correctly categorized 88% ($M \pm SD$: .88 ± .006) of the new tested face pairs. A paired samples *t*-test was performed in order to test for significant differences between neutral-happy and neutral-fearful face pairs (Fig. 11). We found that the network was better able to discriminate neutral from emotional faces when the emotional face was happy ($M \pm SD$: .89 ± .006) than when it was fearful ($M \pm SD$: .87 ± .005; $t(1,49) = 2.55$, $p = .014$, $d = 0.32$).

4.3. Discussion

We used an artificial neural network as a tool to quantify the perceptual differences between neutral and happy compared to neutral and fearful faces. In a similar way to Experiment 2, the neural network reproduced the access to the information from the two hemifields and had to decide which side the emotional face was on based on the perceptual features provided by each hemifield. It is important to clarify that the aim was not to simulate the processing of facial expression perception in the brain, but rather to compare the results obtained from participants and those from an artificial neural network. Overall, results showed that, with an average accuracy of 88%, the network performed better than the participants (on average 65–70%). This can be explained by the fact that saccadic eye movements can be elicited in a bottom-up fashion and thus do not always follow participants' top-down goals. For example, in one and the same task, participants made more errors with saccadic compared to manual responses (Bannerman et al., 2009). Therefore, the fact that we used a saccadic choice rather

than a manual response task in Experiment 2 could have led to a higher rate of erroneous responses.

In line with our hypothesis, we found that the network performances were better when the discrimination involved a happy than a fearful face. This implies that, at a purely perceptual level, fearful and neutral faces may be more similar than happy and neutral faces. Consequently, the fact that participants performed better on happy faces in Experiment 2 might be explicable in terms of perceptual factors. Although we showed that happy faces have a perceptual advantage over fearful faces, our model does not allow us to reject the possibility of a contribution of emotion to the happy face advantage in Experiment 2. For example, we cannot exclude the possibility that happy faces might be prioritized due to their positive valence or because they are encountered more often in everyday life (Bond & Siddle, 1996; Leppänen & Hietanen, 2004).

5. General discussion

The purpose of this study was to assess the impact of facial expressions on selection processes. In the first experiment, we wanted to test the impact of facial expressions on very fast face detection. As previously observed, face targets elicited very fast and accurate saccadic responses, whereas participants took longer and made more errors when required to saccade toward vehicle targets. With regard to the effect of facial expressions, we found that they did not influence performances, thus suggesting that emotional faces, whether happy or fearful, are not automatically (i.e., quickly and nonintentionally) prioritized over neutral faces. Nevertheless, saccade endpoints were modulated by facial expressions. Saccades landed lower when the face was happy than when it was fearful or neutral and also when the face was fearful rather than neutral. In the second experiment, we directly tested the detection of neutral and emotional faces. Emotional faces elicited faster and more accurate responses than neutral ones. Also, accuracy was higher when the emotional face was happy, and saccades landed lower for happy than for fearful face targets. We can note that latencies in Experiment 2 were higher than in Experiment 1 (with mean latencies around 250 ms in Experiment 2 and 170 ms in Experiment 1), suggesting that faces are detected before expressions are explicitly decoded.

5.1. A prioritization of emotional faces?

While emotions modulated performances in the second experiment, emotional faces did not facilitate face detection in the first experiment even though some visual features seemed to differentially attract the gaze depending on facial expression. This observation runs contrary to the idea that attention is reflexively captured by emotional, and especially threatening, events due to evolutionary needs (e.g., Öhman, 2005; Öhman et al., 2001). Nevertheless, similar results (i.e., a fast oculomotor capture by faces irrespective of their expressions) were obtained in an earlier study. Indeed, Devue and Grimshaw (2017) tested the automatic prioritization of nontask-relevant emotional faces in a task in which faces were known to attract the gaze (Devue, Belopolsky, & Theeuwes, 2012). A circular array of colored dots was displayed on the screen and participants had to make a saccade toward a color singleton. Pictures of

different irrelevant objects (including a neutral or an angry face) were displayed in a concentric circle inside the dot array. The authors found that irrelevant faces attracted the gaze more than other objects but that this occurred irrespective of the expression. In their study, saccades toward faces were elicited quickly, with mean saccade latencies around 200 ms. In fact, this lack of modulation could be explained by a ceiling effect (Mulckhuyse, 2018). Indeed, one might consider that the visual system has evolved so that faces, which are more likely to be socially relevant than other objects, can be detected very rapidly based on low-level features (Baron-Cohen, 1995; Haxby, Hoffman, & Gobbini, 2000; Leopold & Rhodes, 2010), or that visual features have evolved to be easily detected (Emery, 2000; Kobayashi & Kohshima, 1997; Lacruz et al., 2019; Wu, Bischof, & Kingstone, 2013). Whether this detection is based on isolated features (i.e., the eyes; Kauffmann et al., 2021; Lewis & Edmonds, 2003) or AS information (Crouzet & Thorpe, 2011; Honey et al., 2008), it may be insufficient to decode expressions. The goal might, therefore, be to direct the face into central vision and then proceed to further investigations in order to extract more features, such as the emotional expression.

Also, in Experiment 1, facial expressions were not task-relevant. However, we suggest that this is not sufficient to explain the lack of emotional modulation. Indeed, there are studies that have used tasks in which expressions were not relevant and which have, nevertheless, found that they modulated performances. This is the case, for example, of probe categorization tasks in which an emotional face is briefly presented as a prime followed by an emotional probe word or a visual scene. The participants' task is to ignore the face (which is task-irrelevant) and judge the probe as pleasant or unpleasant. Classically, reaction times are faster when the prime and probe are affectively congruent, for example, for positive words following an expression of happiness (Aguado, Garcia-Gutierrez, Castañeda, & Saugar, 2007; Lipp et al., 2009; McLellan, Johnston, Dalrymple-Alford, & Porter, 2010; Sassi, Campoy, Castillo, Inuggi, & Fuentes, 2014). Another example can be found in gender categorization tasks. It has been shown that even if expressions are task-irrelevant, they can still interfere with gender perception. For example, it has been shown that a happy or a fearful expression biases discrimination toward females (Hess, Adams, Grammer, & Kleck, 2009) or, similarly, that gender implicitly interferes with the recognition of emotional expressions (Villepoux, Vermeulen, Niedenthal, & Mermillod, 2015). It should be noted that these studies used manual responses, causing reaction times to be relatively high (about 600 ms). If the decoding of facial expressions only occurs after 180–200 ms (Kulke, 2019; Schyns et al., 2009), it is more likely that the lack of emotional prioritization in Experiment 1 was due to the short time window in which saccades were elicited, and that the very fast gaze capture by faces may have preceded emotional facilitation.

In Experiment 2, emotional faces were detected faster and attracted the gaze more often than neutral faces. This confirms that emotional faces, when decoded, can facilitate the orienting of attention. Nevertheless, there is still some doubt as to whether this process is supported by the interpretation of the emotional content. For example, it is possible that when looking for neutral or emotional faces, participants choose to check for emotional faces first, for example, by looking for an open mouth, which is the most salient feature (Calvo & Nummenmaa, 2008; Horstmann et al., 2012; Stuit et al., 2021). Such a strategy would prioritize

emotional faces, not because they are meaningful, but because they have configurations that make them easier to see (because happy and most fearful faces do have an open mouth). In support of this idea, one study which compared emotion detection and emotion categorization suggests that visible teeth are particularly useful for emotion detection (Sweeny, Suzuki, Grabowecky, & Paller, 2013). In Experiment 2, we also found that happy faces led to higher response accuracy than fearful faces. As suggested by neural computations, this last result could be explained by the fact that happy and neutral faces are more perceptually different than fearful and neutral faces (Mermillod et al., 2009).

Moreover, we found in an additional analysis (see Appendix B) that saccades toward faces in Experiment 1 were elicited faster when the face was presented on the left (i.e., projected in the right hemisphere) compared to the right side of the screen (i.e., projected in the left hemisphere). This result agrees with previous papers using a saccadic choice task with faces and vehicles (Crouzet et al., 2010; Guyader et al., 2017). It is also consistent with theories on the cortical lateralization of face processing, which argue in favor of right hemisphere specialization (Carlei, Framorando, Burra, & Kerzel, 2017; Ellis & Young, 1983). Interestingly, saccade endpoints were also lower for targets presented on the left side of the screen. This could also be the result of an enhanced processing of faces in the left hemifield. In Experiment 2, the greater accuracy for emotional than neutral faces was only significant for targets presented on the left side of the screen. This result is consistent with one hypothesis concerning the lateralization of emotional processing, which argues in favor of right hemisphere specialization (Demaree, Everhart, Youngstrom, & Harrison, 2005).

Even if such results are not particularly suitable for inferring neurophysiological implications, they can, nevertheless, be considered in the light of recent neurophysiological data. Indeed, there is evidence that face information can be processed via a subcortical pathway connecting the amygdala, superior colliculus, and pulvinar (Johnson, 2005). Even though some authors believe that this pathway is modulated by facial expressions (Bayle & Taylor, 2010; Bayle, Henaff, & Krolak-Salmon, 2009; LeDoux, 2000; Méndez-Bértolo et al., 2016; Vuilleumier, Armony, Driver, & Dolan, 2003), other studies suggest that it could be face-specific and independent of the expressed emotion (Fitzgerald et al., 2006; Garvert et al., 2014; Johnson, 2005; McFadyen et al., 2017). For example, McFadyen et al. used magnetoencephalography and dynamic causal modeling of participants making gender judgments of neutral and fearful faces to identify the underlying neural networks most likely to carry information to the amygdala. They demonstrated that the most likely subcortical network consisted of a pulvinar–amygdala connection that was not influenced by facial expressions. These results, therefore, suggest that the emotional content of visual stimuli may not necessarily be the key for entry into the subcortical pathway, whereas a more stereotypical face pattern would be. Overall, even if we cannot affirm that face detection in Experiment 1 was supported by such a pathway (e.g., it could also be supported by the ventral pathway; Crouzet et al., 2010), our results would be consistent with this view.

Finally, we can wonder what would have happened if we had used dynamic stimuli. Indeed, some studies have suggested that the use of dynamic stimuli enhances the processing of fearful faces, and, one functional magnetic resonance imaging study identified enhanced neural activity in response to dynamic fearful but not happy faces (Sato, Kochiyama, Yoshikawa,

Naito, & Matsumura, 2004). In addition, an EEG study has shown that dynamic threatening stimuli (e.g., spiders) elicited higher P1 activity than static stimuli or dynamic nonthreatening stimuli (Hinojosa, Carretié, Valcárcel, Méndez-Bértolo, & Pozo, 2009). These findings suggest that motion provides additional saliency to threatening stimuli and thus facilitates their detection. However, there are still only a few studies on this topic and the results are sometimes contradictory. For example, another study found no differences between the responses to dynamic neutral, happy, disgusted, and fearful facial expressions (Van der Gaag, Minderaa, & Keysers, 2007). It is also likely that the results will be similar for both dynamic and static expressions. Studies using static stimuli similar to those that we used in our study (i.e., basic emotions, with the peak frame of the emotion) have reported a lack of a dynamic advantage for expression recognition (Fiorentini & Viviani, 2011; Gold et al., 2013; see for reviews Dobs, Bülthoff, & Schultz, 2018; Kätsyri, 2006). Moreover, saccades are likely to be elicited at the very beginning of the video sequence, in particular in our first experiment.

5.2. Saccade endpoints as a reflection of the distribution of attention within the face?

In both experiments, we found that saccades landed lower when the target was a happy face than when it was a fearful one. This observation was expected given the visual saliency and diagnosticity of the mouth in happy faces (Calvo & Nummenmaa, 2008; Smith et al., 2005). However, saccades also landed lower when the target was a fearful face than when it was a neutral one in Experiment 1. This can be explained by the fact that in fearful faces, both the eyes and mouth attract attention. Indeed, in some fearful faces, the mouth is open and this can shift attention toward it more than neutral faces, in which the mouth is always closed. In line with this view, some studies found that both the eyes and mouth play a critical role in the recognition of fear (Eisenbarth & Alpers, 2011). Given the results of the first experiment, we could have expected that saccades would be lower in the emotional than in the neutral task in Experiment 2. This effect was only marginally significant for happy faces and we suggest that this is due to the small sample size. Nevertheless, perceptual saliency may not be the only factor that can influence the way attention is distributed within the face.

As attested by the heat maps, endpoints were generally located around the eyes in the first experiment, and more around the nose for the second experiment. Such differences between the first and second experiment can be explained by the different tasks. It is likely that when participants have to process facial expressions, their gaze is naturally more oriented toward the mouth. It is also important to underline that because the eyes were located in the center of the image in this study, it is possible that saccades tended to land around the eyes because of a center of gravity bias (Bindemann, 2010; Parkhurst & Niebur, 2003; Tatler, 2007; Tseng, Carmi, Cameron, Munoz, & Itti, 2009). Furthermore, even if we assume that the positions of the eyes and mouth were the same for all images, some slight differences might, nevertheless, have occurred, again due to the different, emotion-related shapes of the eyes and mouth. To explore such differences, we performed an additional analysis (see Appendix A) demonstrating that the positions of the vertical midline of the eyes and mouth were not uniform across expressions. Indeed, the mouth position was higher on average in the image of happy than those of neutral or fearful faces, as well as in the images of neutral faces than fearful ones.

Also, the mean eye position was higher for fearful than for happy or neutral faces, and for happy than for neutral faces. Therefore, the difference between saccade endpoints on happy and fearful faces (i.e., lower endpoints on happy faces) could be explained by the fact that the eyes are lower in happy faces. However, this explanation on its own would not explain the difference between neutral and fearful faces in Experiment 1 (lower endpoints on fearful faces) since, based on the eye position, we would have expected lower endpoints for neutral faces.

Overall, we assume that attention is shared between multiple locations during saccade programming and that saccade endpoints can reflect the interactions between multiple loci of attention. In this context, when a saccade is executed, the endpoint would reflect the allocation of attention within the face but would not necessarily indicate the exact location that has captured most of the attention. For example, when a saccade has to be executed toward a target (e.g., a dark gray ring) in the presence of a close distractor (e.g., a dark ring), the saccade lands somewhere in-between the target and the distractor. Van der Stigchel and de Vries (2015) have suggested that attention is directed toward the target and distractor location rather than at the intermediate location (where the saccade lands). This is consistent with neurophysiological recordings in such configurations which have shown that peaks of activity in the superior colliculus are focused on cells coding for target and distractor locations (Edelman & Keller, 1998). This phenomenon is often referred to as global effect or saccade averaging, and has, to our knowledge, only been studied using very basic stimuli, such as circles with line drawings (Findlay, 1982; Van der Stigchel & Nijboer, 2013). Given that we compared emotional faces, we suggest that attention is mostly shared between face parts which are of diagnostic relevance for expression decoding, which have been shown to be the eyes or mouth (Eisenbarth & Alpers, 2011; Smith et al., 2005; Wegrzyn et al., 2017). We assume that if the endpoint is lower in one condition, it means that the mouth attracted most attention during saccade programming. Following this view, even though saccade endpoints were located around the nose in the second experiment, this does not necessarily mean that attention was directed there.

6. Conclusions

The present study confirms previous results showing that face stimuli can be detected very efficiently and provides new insights concerning the interaction between facial expression processing and the oculomotor system. Experiment 1 shows that very fast face detection was not modulated by facial expressions, suggesting that emotional faces, whether fearful or happy, are not automatically prioritized over neutral ones. Experiment 2 showed that emotional faces are detected better than neutral ones in a task in which participants are explicitly asked to process expressions. We suggest that the lack of emotional prioritization in Experiment 1 is due to the short time window in which saccades were elicited (with mean saccade latencies around 170 ms). Indeed, as suggested by some previous studies (Devue et al., 2012; Kulke, 2019; Schyns et al., 2009), fast face detection may occur before expressions are decoded. Experiment 2 also found that it was easier to discriminate between neutral and

emotional faces in the case of happy rather than fearful faces. Using computational modeling, we showed that this can at least be partly explained by perceptual factors, as the performances of a neural network were also better with happy faces. Finally, an analysis of saccade endpoints revealed a modulation by facial expressions, even for saccades that were elicited very quickly. We suggest that salient local face features, like the mouth, can automatically shift attention toward themselves and all the more so when they are task-relevant.

Acknowledgments

This work was supported by NeuroCoG IDEX UGA in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02). This work has been partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003) to Martial Mermillod.

Conflict of interest

The authors declare no conflict of interest.

References

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*(3), 165–178. https://doi.org/10.1038/nrn1056

Aguado, L., Garcia-Gutierrez, A., Castañeda, E., & Saugar, C. (2007). Effects of prime task on affective priming by facial expressions of emotion. *Spanish Journal of Psychology*, *10*(2), 209–217. https://doi.org/10.1017/S1138741600006478

Bannerman, R. L., Milders, M., & Sahraie, A. (2009). Processing emotional stimuli: Comparison of saccadic and manual choice-reaction times. *Cognition & Emotion*, *23*(5), 930–954. https://doi.org/10.1080/02699930802243303

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 41–59). Lawrence Erlbaum.

Bayle, D. J., Henaff, M.-A., & Krolak-Salmon, P. (2009). Unconsciously perceived fear in peripheral vision alerts the limbic system: A MEG study. *PLoS One*, *4*(12), e8207. https://doi.org/10.1371/journal.pone.0008207

Bayle, D. J., & Taylor, M. J. (2010). Attention inhibition of early cortical activation to fearful faces. *Brain Research*, *1313*, 113–123. https://doi.org/10.1016/j.brainres.2009.11.060

Becker, D. V., Anderson, U. S., Mortensen, C. R., Neufeld, S. L., & Neel, R. (2011). The face in the crowd effect unconfounded: Happy faces, not angry faces, are more efficiently detected in single- and multiple-target visual search tasks. *Journal of Experimental Psychology: General*, *140*(4), 637–659. https://doi.org/10.1037/a0024060

Belopolsky, A. V. (2015). Common priority map for selection history, reward and emotion in the oculomotor system. *Perception*, *44*(8–9), 920–933. https://doi.org/10.1177/0301006615596866

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, *50*(23), 2577–2587. https://doi.org/10.1016/j.visres.2010.08.016

Bisley, J. W., & Mirpour, K. (2019). The neural instantiation of a priority map. *Current Opinion in Psychology*, *29*, 108–112. https://doi.org/10.1016/j.copsyc.2019.01.002

Bond, N. W., & Siddle, D. A. T. (1996). The preparedness account of social phobia: Some data and alternative explanations. In R. M. Rapee (Ed.), *Current controversies in the anxiety disorders* (pp. 291–316). London: Guilford Press.

Boucart, M., Lenoble, Q., Quettelart, J., Szaffarczyk, S., Despretz, P., & Thorpe, S. J. (2016). Finding faces, animals, and vehicles in far peripheral vision. *Journal of Vision*, *16*(2), 10–10.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109–115. https://doi.org/10.3758/BRM.40.1.109

Calvo, M. G., & Marrero, H. (2009). Visual search of emotional faces: The role of affective content and featural distinctiveness. *Cognition & Emotion*, *23*(4), 782–806. https://doi.org/10.1080/02699930802151654

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: Salient physical features guide effective visual search. *Journal of Experimental Psychology: General*, *137*(3), 471–494. https://doi.org/10.1037/a0012771

Calvo, M. G., & Nummenmaa, L. (2009). Eye-movement assessment of the time course in facial expression recognition: Neurophysiological implications. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(4), 398–411. https://doi.org/10.3758/CABN.9.4.398

Calvo, M. G., & Nummenmaa, L. (2011). Time course of discrimination between emotional facial expressions: The role of visual saliency. *Vision Research*, *51*(15), 1751–1759. https://doi.org/10.1016/j.visres.2011.06.001

Calvo, M. G., & Nummenmaa, L. (2015). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, *30*(6), 1081–1106. https://doi.org/10.1080/02699931.2015.1049124

Carlei, C., Framorando, D., Burra, N., & Kerzel, D. (2017). Face processing is enhanced in the left and upper visual hemi-fields. *Visual Cognition*, *25*(7–8), 749–761.

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, *9*(12), 10–10. https://doi.org/10.1167/9.12.10

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 5. https://doi.org/10.1167/14.8.5

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 16–16. https://doi.org/10.1167/10.4.16

Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology*, *2*, 342.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience*, *14*(8), 1158–1173. https://doi.org/10.1162/089892902760807177

Demaree, H. A., Everhart, D. E., Youngstrom, E. A., & Harrison, D. W. (2005). Brain lateralization of emotional processing: Historical roots and a future incorporating "dominance". *Behavioral and Cognitive Neuroscience Reviews*, *4*(1), 3–20.

Devue, C., Belopolsky, A. V., & Theeuwes, J. (2012). Oculomotor guidance and capture by irrelevant faces. *PLoS One*, *7*(4), e34598. https://doi.org/10.1371/journal.pone.0034598

Devue, C., & Grimshaw, G. M. (2017). Faces are special, but facial expressions aren't: Insights from an oculomotor capture paradigm. *Attention, Perception, & Psychophysics*, *79*(5), 1438–1452. https://doi.org/10.3758/s13414-017-1313-x

Dobs, K., Bülthoff, I., & Schultz, J. (2018). Use and usefulness of dynamic face stimuli for face perception studies—A review of behavioral findings and methodology. *Frontiers in Psychology*, *9*, 1355.

Edelman, J. A., & Keller, E. L. (1998). Dependence on target configuration of express saccade-related activity in the primate superior colliculus. *Journal of Neurophysiology*, *80*(3), 1407–1426. https://doi.org/10.1152/jn.1998.80.3.1407

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion*, *11*(4), 860–865. https://doi.org/10.1037/a0022758

Ellis, H. D., & Young, A. (1983). The role of the right hemisphere in face perception. In A. Young, *Functions of the right cerebral hemisphere* (pp. 33–64).

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581–604.

Fecteau, J., & Munoz, D. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Findlay, J. M. (1982). Global visual processing for saccadic eye movements. *Vision Research*, *22*(8), 1033–1045. https://doi.org/10.1016/0042-6989(82)90040-2

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision*, *11*(3), 17–17.

Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *Neuroimage*, *30*(4), 1441–1448. https://doi.org/10.1016/j.neuroimage.2005.11.003

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, *117*(3), 319–331. https://doi.org/10.1016/j.cognition.2010.09.003

Fox, E., & Damjanovic, L. (2006). The eyes are sufficient to produce a threat superiority effect. *Emotion*, *6*(3), 534–539. https://doi.org/10.1037/1528-3542.6.3.534

Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological Bulletin*, *134*(5), 662–676. https://doi.org/10.1037/0033-2909.134.5.662

Garvert, M. M., Friston, K. J., Dolan, R. J., & Garrido, M. I. (2014). Subcortical amygdala pathways enable rapid face processing. *Neuroimage*, *102*, 309–316. https://doi.org/10.1016/j.neuroimage.2014.07.047

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., … Srinath, A. (2013). The efficiency of dynamic and static facial expression recognition. *Journal of Vision*, *13*(5), 23–23.

Gu, X., Hoijtink, H., Mulder, J., & Rosseel, Y. (2019). Bain: A program for Bayesian testing of order constrained hypotheses in structural equation models. *Journal of Statistical Computation and Simulation*, *89*(8), 1526–1553.

Guyader, N., Chauvin, A., Boucart, M., & Peyrin, C. (2017). Do low spatial frequencies explain the extremely fast saccades towards human faces? *Vision Research*, *133*, 100–111. https://doi.org/10.1016/j.visres.2016.12.019

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

Hess, U., Adams, R. B., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: Are angry women more like men? *Journal of Vision*, *9*(12), 19–19. https://doi.org/10.1167/9.12.19

Hinojosa, J. A., Carretié, L., Valcárcel, M. A., Méndez-Bértolo, C., & Pozo, M. A. (2009). Electrophysiological differences in the processing of affective information in words and pictures. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(2), 173–189.

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, *24*(5), 539

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *Journal of Vision*, *8*(12), 9–9.

Horstmann, G., Lipp, O., & Becker, S. (2012). Of toothy grins and angry snarls—Open mouth displays contribute to efficiency gains in search for emotional faces. *Journal of Vision*, *12*(5), 7. https://doi.org/10.1167/12.5.7.

Hoskin, T. (2012). Parametric and nonparametric: Demystifying the terms. *Mayo Clinic*, *5*, 1–5.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Humphrey, K., Underwood, G., & Lambert, T. (2012). Salience of the lambs: A test of the saliency map hypothesis with pictures of emotive objects. *Journal of Vision*, *12*(1), 22–22. https://doi.org/10.1167/12.1.22

Hunt, A. R., Cooper, R. M., Hungr, C., & Kingstone, A. (2007). The effect of emotional faces on eye movements and attention. *Visual Cognition*, *15*(5), 513–531. https://doi.org/10.1080/13506280600843346

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Jeffreys, H. (1998). *The theory of probability*. Oxford: OUP.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, *6*(10), 766–774. https://doi.org/10.1038/nrn1766

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*(6), 1233–1258. https://doi.org/10.1152/jn.1987.58.6.1233

Juth, P., Lundqvist, D., Karlsson, A., & Ohman, A. (2005). Looking for foes and friends: Perceptual and emotional factors when finding a face in the crowd. *Emotion*, *5*(4), 379–395.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kätsyri, J. (2006). Human recognition of basic emotions from posed and animated dynamic facial expressions.

Kauffmann, L., Khazaz, S., Peyrin, C., & Guyader, N. (2021). Isolated face features are sufficient to elicit ultra-rapid and involuntary orienting responses toward faces. *Journal of Vision*, *21*(2), 4–4.

Kauffmann, L., Peyrin, C., Chauvin, A., Entzmann, L., Breuil, C., & Guyader, N. (2019). Face perception influences the programming of eye movements. *Scientific Reports*, *9*(1). 1–14. https://doi.org/10.1038/s41598-018-36510-0

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. ArXiv:1412.6980 [Cs]. Retrieved from http://arxiv.org/abs/1412.6980

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776. https://doi.org/10.1016/j.visres.2005.10.002

Klink, P. C., Jentgens, P., & Lorteije, J. A. M. (2014). Priority maps explain the roles of value, attention, and salience in goal-oriented behavior. *Journal of Neuroscience*, *34*(42), 13867–13869. https://doi.org/10.1523/JNEUROSCI.3249-14.2014

Kloth, N., Itier, R. J., & Schweinberger, S. R. (2013). Combined effects of inversion and feature removal on N170 responses elicited by faces and car fronts. *Brain and Cognition*, *81*(3), 321–328. https://doi.org/10.1016/j.bandc.2013.01.002

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*(6635), 767–768.

Kulke, L. (2019). Neural mechanisms of overt attention shifts to emotional faces. *Neuroscience*, *418*, 59–68.

Lacruz, R. S., Stringer, C. B., Kimbel, W. H., Wood, B., Harvati, K., O'Higgins, P., … Arsuaga, J. L. (2019). The evolutionary history of the human face. *Nature Ecology & Evolution*, *3*(5), 726–736.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184.

Leopold, D. A., & Rhodes, G. (2010). A comparative view of face perception. *Journal of Comparative Psychology*, *124*(3), 233.

Leppänen, J. M., & Hietanen, J. K. (2004). Emotionally positive facial expressions are processed faster than negative facial expressions, but why? *Psychological Research*, *69*, 2229.

Lewis, M. B., & Edmonds, A. J. (2003). Face detection: Mapping human performance. *Perception*, *32*(8), 903–920.

Li, R., & Cottrell, G. (2012). A new angle on the EMPATH model: Spatial frequency orientation in recognition of facial expressions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Lilliefors, H. W. (1967). On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, *62*(318), 399–402. https://doi.org/10.1080/01621459.1967.10482916

Lipp, O. V., Price, S. M., & Tellegen, C. L. (2009). No effect of inversion on attentional and affective processing of facial expressions. *Emotion*, *9*(2), 248–259. https://doi.org/10.1037/a0014715

Lundqvist, D., Flykt, A., & Ohman, A. (1998). *The Karolinska directed emotional faces (KDEF)*.

Marat, S., Ho Phuoc, T., Granjon, L., Guyader, N., Pellerin, D., & Guérin-Dugué, A. (2009). Modelling spatio-temporal saliency to predict gaze direction for short videos. *International Journal of Computer Vision*, *82*(3), 231–243. https://doi.org/10.1007/s11263-009-0215-3

Martin, J. G., Davis, C. E., Riesenhuber, M., & Thorpe, S. J. (2018). Zapping 500 faces in less than 100 seconds: Evidence for extremely fast and sustained continuous visual search. *Scientific Reports*, *8*(1), 1–12.

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, *46*(1), 38–41. https://doi.org/10.1016/j.bjoms.2007.09.002

McFadyen, J., Mermillod, M., Mattingley, J. B., Halász, V., & Garrido, M. I. (2017). A rapid subcortical amygdala route for faces irrespective of spatial frequency and emotion. *Journal of Neuroscience*, *37*(14), 3864–3874. https://doi.org/10.1523/JNEUROSCI.3525-16.2017

McLellan, T., Johnston, L., Dalrymple-Alford, J., & Porter, R. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cognition & Emotion*, *24*(8), 1277–1292. https://doi.org/10.1080/02699930903306181

Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y. H., … Strange, B. A. (2016). A fast pathway for fear in human amygdala. *Nature Neuroscience*, *19*(8), 1041–1049. https://doi.org/10.1038/nn.4324

Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation. *Neurocomputing*, *73*(13–15), 2522–2531. https://doi.org/10.1016/j.neucom.2010.06.002

Mermillod, M., Bourrier, Y., David, E., Kauffmann, L., Chauvin, A., Guyader, N., … Peyrin, C. (2019). The importance of recurrent top-down synaptic connections for the anticipation of dynamic emotions. *Neural Networks*, *109*, 19–30. https://doi.org/10.1016/j.neunet.2018.09.007

Mermillod, M., Vermeulen, N., Lundqvist, D., & Niedenthal, P. M. (2009). Neural computation as a tool to differentiate perceptual from emotional processes: The case of anger superiority effect. *Cognition*, *110*(3), 346–357. https://doi.org/10.1016/j.cognition.2008.11.009

Morris, J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain*, *121*(1), 47–57. https://doi.org/10.1093/brain/121.1.47

Mulckhuyse, M. (2018). The influence of emotional stimuli on the oculomotor system: A review of the literature. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(3), 411–425. https://doi.org/10.3758/s13415-018-0590-8

Niu, Y., Todd, R. M., & Anderson, A. (2012). Affective salience can reverse the effects of stimulus-driven salience on eye movements in complex scenes. *Frontiers in Psychology*. *3*, 336. https://doi.org/10.3389/fpsyg.2012.00336

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology*, *30*(10), 953–958. https://doi.org/10.1016/j.psyneuen.2005.03.019

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*(3), 381–396. https://doi.org/10.1037/0022-3514.80.3.381

Parkhurst, D., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, *16*(2), 125–154. https://doi.org/10.1163/15685680360511645

Racine J. (2012). RStudio: A Platform-Independent IDE for R and Sweave. *Journal of Applied Econometrics*, *27*(1), 167–172. http://doi.org/10.1002/jae.1278

R Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Sassi, F., Campoy, G., Castillo, A., Inuggi, A., & Fuentes, L. J. (2014). Task difficulty and response complexity modulate affective priming by emotional facial expressions. *Quarterly Journal of Experimental Psychology*, *67*(5), 861–871. https://doi.org/10.1080/17470218.2013.836233

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: An fMRI study. *Cognitive Brain Research*, *20*(1), 81–91.

Schubö, A., Gendolla, G. H. E., Meinecke, C., & Abele, A. E. (2006). Detecting emotional faces and features in a visual search paradigm: Are faces special? *Emotion*, *6*(2), 246–256. https://doi.org/10.1037/1528-3542.6.2.246

Schyns, P. G., Petro, L. S., & Smith, M. L. (2009). Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: Behavioral and brain evidence. *PLoS One*, *4*(5), e5625. https://doi.org/10.1371/journal.pone.0005625

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*(3), 184–189. https://doi.org/10.1111/j.0956-7976.2005.00801

Stuit, S. M., Kootstra, T. M., Terburg, D., van den Boomen, C., van der Smagt, M. J., Kenemans, J. L., & Van der Stigchel, S. (2021). The image features of emotional faces that predict the initial eye movement to a face. *Scientific Reports*, *11*(1), 1–14.

Sweeny, T. D., Suzuki, S., Grabowecky, M., & Paller, K. A. (2013). Detecting and categorizing fleeting emotions in faces. *Emotion*, *13*(1), 76–91. https://doi.org/10.1037/a0029193

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, *11*(10), 697–709. https://doi.org/10.1038/nrn2889

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 4. https://doi.org/10.1167/7.14.4

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology*, *29*, 97–101. https://doi.org/10.1016/j.copsyc.2018.12.024

Tipples, J., Atkinson, A. P., & Young, A. W. (2002). The eyebrow frown: A salient social signal. *Emotion*, *2*(3), 288–296. https://doi.org/10.1037/1528-3542.2.3.288

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Tseng, P. H., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, *9*(7), 4–4. https://doi.org/10.1167/9.7.4

Van der Gaag, C., Minderaa, R. B., & Keysers, C. (2007). The BOLD signal in the amygdala does not differentiate between dynamic facial expressions. *Social Cognitive and Affective Neuroscience*, *2*(2), 93–103.

Van der Stigchel, S., & de Vries, J. P. (2015). There is no attentional global effect: Attentional shifts are independent of the saccade endpoint. *Journal of Vision*, *15*(15), 17. https://doi.org/10.1167/15.15.17

Van der Stigchel, S., & Nijboer, T. C. W. (2013). How global is the global effect? The spatial characteristics of saccade averaging. *Vision Research*, *84*, 6–15. https://doi.org/10.1016/j.visres.2013.03.006

Villepoux, A., Vermeulen, N., Niedenthal, P., & Mermillod, M. (2015). Evidence of fast and automatic gender bias in affective priming. *Journal of Cognitive Psychology*, *27*(3), 301–309

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, *6*(6), 624–631. https://doi.org/10.1038/nn1057

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PLoS One*, *12*(5), e0177239. https://doi.org/10.1371/journal.pone.0177239

Wu, D. W. L., Bischof, W. F., & Kingstone, A. (2014). Natural gaze signaling in a social context. *Evolution and Human Behavior*, *35*(3), 211–218.

2.3 Expérience complémentaire : Visage masculin vs Visage féminin

2.3.1 Introduction

Dans l'Article 1, nous avons montré que les visages émotionnels peuvent être détectés et attirer le regard plus efficacement que les visages neutres (ce qui se traduit par des proportions de saccades correctes plus élevées et des latences plus faibles). Néanmoins, cet effet n'était pas automatique, puisqu'il n'a été observé que dans l'Expérience 2, lorsque la tâche était de faire une saccade vers le visage émotionnel ou le visage neutre. Dans l'Expérience 1, lorsque la tâche était de faire une saccade vers le visage (avec un véhicule comme distracteur), les visages émotionnels n'attiraient pas plus vite ou plus souvent le regard que les visages neutres. En discussion de l'Article 1, nous suggérons que cette absence d'effet observé dans l'Expérience 1 (Visage vs Véhicule) soit liée aux latences des saccades, qui sont très faibles lorsqu'il s'agit de détecter un visage. À ce niveau dans le décours temporel de la perception visuelle, les visages seraient détectés de manière grossière, et ce traitement grossier ne suffirait pas pour discriminer les émotions. Néanmoins, les différences observées entre les résultats de l'Expérience 1 (Visage vs Véhicule) et de l'Expérience 2 (Visage émotionnel vs. Visage neutre) pourraient aussi s'expliquer par des différences en termes d'orientation de l'attention. Dans l'Expérience 2, l'attention des participants est orientée vers les émotions du visage, qui sont pertinentes pour la tâche ; on étudie alors leur traitement explicite. Au contraire, dans l'Expérience 1 les émotions du visage ne sont pas pertinentes pour la tâche ; on étudie alors leur traitement implicite.

Plusieurs études suggèrent que l'activité de l'amygdale est indépendante de l'orientation de l'attention. Ainsi, elle différencierait de la même manière les expressions faciales, quelle que soit la tâche des participants (par exemple une tâche de détection de genre, d'identification ou de détection d'émotions ; Anderson et al., 2003 ; Vuilleumier, 2002). D'autres études ont néanmoins mis en évidence une modulation en fonction du type de traitement (implicite ou explicite). Par exemple, dans une étude en MEG, Bayle et al. (2010) ont mis en évidence une discrimination précoce des émotions (90 ms) seulement dans une condition de traitement implicite (l'attention était dirigée vers le traitement de l'identité des visages). Le traitement explicite des émotions induisait néanmoins des activations plus fortes aux alentours de 170 ms. Dans une étude en IRMf, Williams et al. (2005) ont observé une activité de l'amygdale plus importante en condition de traitement explicite pour les visages heureux, et une activité de l'amygdale plus importante en condition de traitement implicite pour les visages apeurés. Pour finir, en utilisant plusieurs expressions faciales différentes (joyeuses, tristes, en colère, apeurées et dégoûtées) Habel et al. (2007) ont observé une activité de l'amygdale supérieure, en moyenne, dans le cas d'un traitement explicite.

Au niveau comportemental, plusieurs études ont utilisé des tâches dans lesquelles les expressions n'étaient pas pertinentes, et ont montré qu'elles modulaient quand même les performances (Bannerman et al., 2012b ; S. I. Becker et al., 2017 ; D'Hondt et al., 2016 ; Nummenmaa et al., 2006). C'est le cas, par exemple des tâches de catégorisation dans

lesquelles un visage émotionnel est brièvement présenté comme amorce, suivi d'un mot ou d'une scène visuelle. La tâche des participants consiste à ignorer le visage (qui n'est donc pas pertinent pour la tâche) et à juger le mot ou la scène comme étant agréable ou désagréable. Classiquement, les temps de réaction sont plus courts lorsque l'amorce et le stimulus à évaluer sont affectivement congruents, par exemple pour des mots positifs suivant une expression joyeuse (Aguado et al., 2007 ; Lipp et al., 2009 ; McLellan et al., 2010 ; Sassi et al., 2014). Plusieurs études ont mis en évidence un traitement privilégié des scènes émotionnelles en utilisant des mesures oculométriques et des tâches dans lesquelles le contenu émotionnel n'était pas pertinent. Par exemple, dans une étude de D'Hondt et al. (2016) un paradigme de choix saccadique avec des paires de scènes présentées à différentes excentricités, l'une ovale et l'autre rectangulaire, a été utilisé. Les participants devaient faire une saccade vers la scène ovale (dans une session) ou rectangulaire (dans une autre session). Les auteurs ont mis en évidence de meilleures performances lorsque la cible était une scène émotionnelle que neutre (D'Hondt et al., 2016 ; cet effet était néanmoins limité à une excentricité de 10°). Dans ce sens, une autre étude a montré que, lorsque des paires de scènes, l'une émotionnelle et l'autre neutre, sont présentées, la probabilité que la première fixation soit sur la scène émotionnelle est plus élevée, même lorsqu'il est demandé aux participants de regarder d'abord le stimulus neutre (Nummenmaa et al., 2006).

Le but de cette expérience était de montrer que, même lorsqu'elles ne sont pas pertinentes pour la tâche, les expressions faciales émotionnelles peuvent capturer l'attention et le regard plus que les expressions neutres. Afin d'optimiser la comparaison avec l'Expérience 2, nous avons repris la même procédure et les mêmes stimuli, en changeant seulement la consigne. Dans l'Expérience 2, chaque essai opposait, non seulement un visage émotionnel et un visage neutre, mais aussi un visage féminin et un visage masculin. Dans cette expérience complémentaire, nous avons demandé aux participants d'orienter leur regard le plus rapidement possible vers le visage masculin dans une session, et vers le visage féminin, dans une autre session. Ainsi, la seule différence par rapport à l'Expérience 2 est qu'ici l'attention n'est plus orientée vers l'expression faciale des visages, mais vers le genre des visages.

Du fait que nous étudions dans cette expérience la perception du genre, il est nécessaire de considérer la littérature concernant les interactions entre la perception des expressions faciales et la perception du genre. Par exemple, une revue de la littérature sur le sujet fait état d'un impact du genre sur la reconnaissance des émotions et vice-versa (Adams Jr et al., 2015). Cette revue souligne le fait que la représentation du genre est généralement associée à des attentes stéréotypées concernant le comportement émotionnel des individus. Par exemple, les expressions joyeuses, apeurées ou tristes sont plus souvent associées aux femmes (perçues comme plus émotives), et les expressions neutres ou en colère sont plus souvent associées aux hommes. Ce constat est notamment appuyé par une étude dans laquelle des visages androgynes sont présentés à des participants, qui doivent les catégoriser selon leur genre. Si les visages avaient une expression qui se rapprochait de la colère, ils étaient plus facilement catégorisés comme masculins, alors que s'ils avaient

des traits qui se rapprochaient de la tristesse ou de la joie, ils étaient plus facilement catégorisés comme féminins (Hess et al., 2009).

Principalement, nous nous attendions à ce que, même dans une tâche dans laquelle l'attention des participants n'est pas orientée vers les expressions faciales, les visages émotionnels attirent plus facilement l'attention que les visages neutres (et donc à des proportions de saccades correctes plus élevées et des latences plus faibles lorsque la cible est émotionnelle plutôt que neutre). De plus, puisque nous avons utilisé des visages émotionnels apeurés ou joyeux, plus facilement associés au genre féminin, nous supposions que l'avantage des visages émotionnels serait plus fort lorsque la cible est un visage féminin. Concernant les différences entre les visages joyeux et apeurés, comme dans l'Article 1, deux hypothèses alternatives ont été dissociées. Les visages joyeux pourraient attirer plus facilement le regard du fait qu'ils sont plus reconnaissables, et les visages apeurés du fait qu'ils sont pertinents pour la survie. Pour finir, nous avons aussi émis l'hypothèse que les participants soient meilleurs pour détecter le genre d'un individu de leur propre groupe. Par exemple, une femme détecterait plus facilement le visage d'une femme, conformément aux résultats d'une précédente étude (Cellerino et al., 2004).

2.3.2 Méthode

2.3.2.1 Participants

Cette étude a fait l'objet d'un pré-enregistrement sur l'*Open Science Framework* (https://osf.io/vynup). Quarante-cinq participants volontaires (vingt-trois femmes ; $M \pm SD$: 20.1 ± 0.55 ; tranche d'âge : 18-30 ans) ont été recrutés pour effectuer une tâche de choix saccadique. Tous les participants avaient une vision normale ou corrigée à la normale (par l'intermédiaire du port de lunettes ou de lentilles), et aucun d'entre eux n'a reporté de maladie psychiatrique ou neurologique passée ou présente. Les étudiants en psychologie à l'Université Grenoble Alpes ont reçu des points pour leur participation à l'expérience. Tous les participants ont donné leur consentement écrit et éclairé avant l'expérience, qui a été réalisée conformément au Code d'éthique de l'Association médicale mondiale (Déclaration d'Helsinki) pour les expériences impliquant des êtres humains.

2.3.2.2 Stimuli et procédure

Les stimuli étaient les mêmes que ceux utilisés dans l'Expérience 2 de l'Article 1. Il y avait donc 180 visages, qui correspondaient à 60 acteurs (30 femmes) dont l'expression était neutre, joyeuse ou apeurée. Les images avaient une taille de 300×300 pixels, correspondant à $11 \times 11°$ d'angle visuel dans la configuration de l'expérience. La procédure était aussi similaire à celle de l'Expérience 2 de l'Article 1, seule la consigne était différente. Toujours, un visage émotionnel et un visage neutre étaient simultanément présentés à l'écran pendant 800 ms, l'un correspondait à un visage féminin, l'autre à un visage masculin. Mais cette fois, il était demandé aux participants d'orienter le plus rapidement possible leur regard vers le visage masculin dans une session et vers le visage féminin dans une autre session.

2.3.2.3 Analyse des données

Les méthodes de prétraitement et d'analyse des données étaient similaires à celles des expériences de l'Article 1. La procédure de prétraitement a entraîné le rejet de 10.9 % du nombre initial d'essais. Les mêmes variables indépendantes, caractérisant la première saccade des participants, ont été analysées dans chaque condition expérimentale. Ainsi, pour chaque participant, nous avons calculé la proportion de saccades correctes, la latence moyenne pour les essais corrects, ainsi que la distance verticale moyenne au centre pour les essais corrects. Une ANOVA mixte a été appliquée sur chacune de ces variables avec la Cible (Visage Masculin, Visage Féminin) et l'Expression Faciale Émotionnelle de la cible (EFE ; Joyeuse, Apeurée, Neutre) comme facteurs intra-sujets, et le Genre du participant (Homme, Femme) comme facteur inter-sujets. Si nécessaire (c'est-à-dire si un effet d'interaction était observé), des tests t à échantillons appariés étaient utilisés pour les comparaisons par paires, et une correction de Bonferroni était appliquée pour corriger le seuil de significativité des comparaisons multiples. Les tailles d'effet ont été estimées en calculant l'êta carré partiel (η_p^2), et un effet était considéré comme significatif si sa valeur p était inférieure au seuil $\alpha = .05$.

2.3.3 Résultats

2.3.3.1 Proportion de saccades correctes

L'ANOVA mixte effectuée sur les proportions de saccades correctes (Figure 2.1-a) a indiqué un effet significatif de l'EFE ($F(2,86) = 15.4$, $p < .001$, $\eta_p^2 = .26$). La proportion de saccades correctes était plus élevée lorsque la cible était joyeuse ($M \pm SD : .75 \pm .092$; $p < .001$) ou apeurée ($M \pm SD : .73 \pm .1$; $p = .013$) que neutre ($M \pm SD : .71 \pm .1$). Elle était également plus élevée lorsque la cible était joyeuse qu'apeurée ($p = .009$). Une interaction significative entre le Genre et la Cible ($F(1,43) = 11.4$, $p = .002$, $\eta_p^2 = .069$) a été observée, ainsi qu'une interaction marginale entre le Genre et l'EFE ($F(2,86) = 3.1$, $p = .051$, $\eta_p^2 = .21$). Cependant, les comparaisons par paires n'ont révélé aucune différence significative [2].

2.3.3.2 Latences

L'ANOVA mixte effectuée sur les latences moyennes des saccades correctes (Figure 2.1-b) n'a indiqué aucun effet significatif.

2. Bien que cela ne soit pas significatif et que nous ne puissions pas conclure sur ces effets d'interaction, nous pouvons supposer que l'interaction entre le Genre et la Cible émerge du fait que la proportion de saccades correctes pour les femmes est plus élevée, en moyenne, lorsque la cible est le visage féminin ($M \pm SD : .75 \pm .1$) que le visage masculin ($M \pm SD : .72 \pm .1$), alors que c'est le contraire pour les hommes ($M \pm SD$ *pour la cible féminine* $: .71 \pm .1$; $M \pm SD$ *pour la cible masculine* $: .75 \pm .09$). Pour l'interaction entre le Genre et l'EFE, elle pourrait émerger du fait que pour les femmes la proportion de saccades correctes pour les visages apeurés ($M \pm SD : .718 \pm .11$) et neutres ($M \pm SD : .715 \pm .1$) est similaire, alors que pour les hommes la proportion de saccades correctes pour les visages apeurés ($M \pm SD : .74 \pm .1$) et joyeux ($M \pm SD : .749 \pm .1$) est similaire.

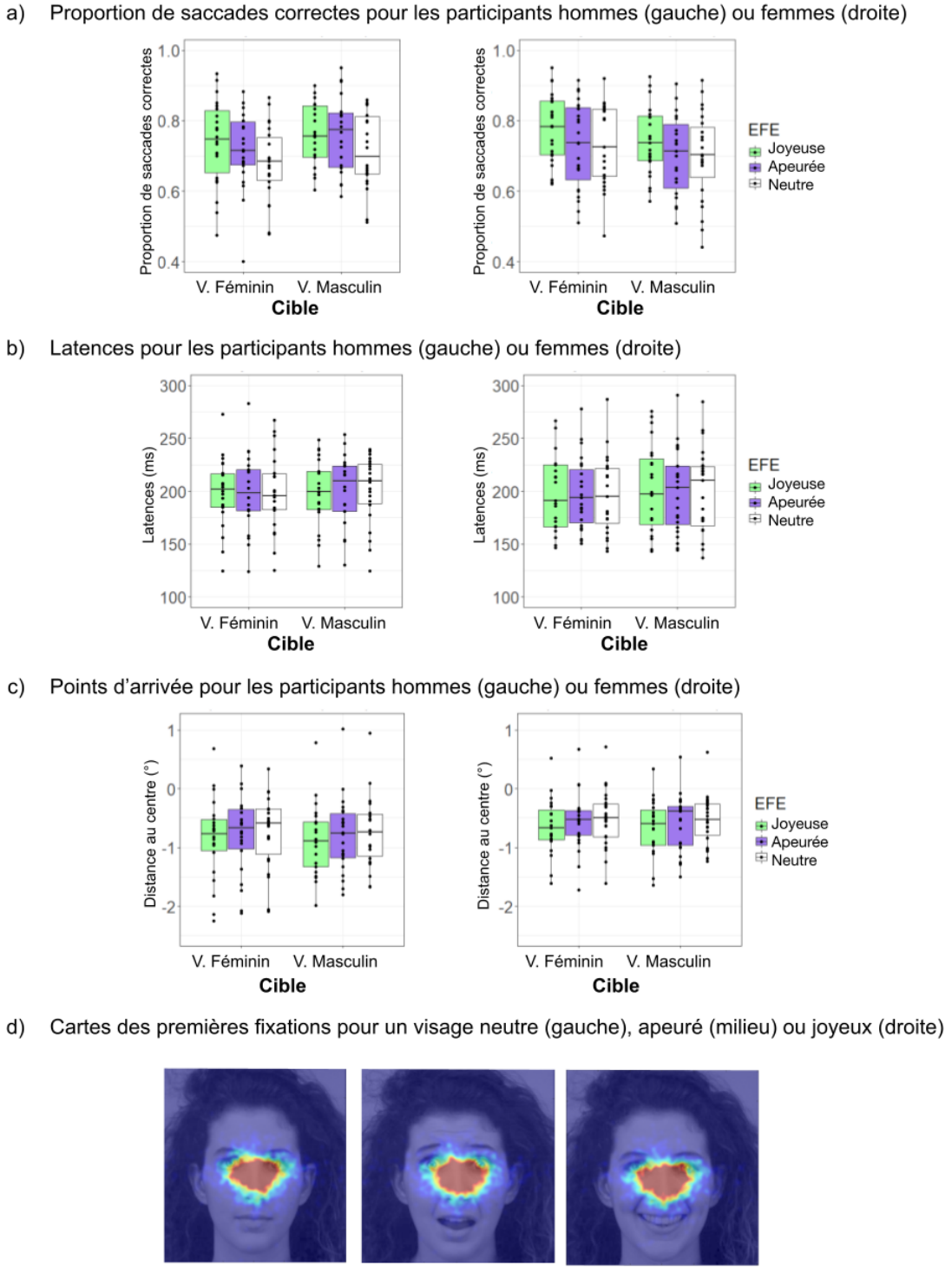

Figure 2.1 – Résultats de l'expérience. (a) Proportion de saccades correctes, (b) latence des saccades correctes et (c) distance verticale par rapport au centre des points d'arrivée des saccades correctes pour les participants hommes (gauche) ou femmes (droite) en fonction de la cible (le visage féminin ou masculin) et de l'expression faciale émotionnelle de la cible (joyeuse, apeurée ou neutre). (d) Visualisation des points d'arrivée des premières saccades en fonction de l'expression faciale émotionnelle de la cible.

2.3.3.3 Points d'arrivée

L'ANOVA mixte effectuée sur la distance verticale entre le centre de l'image et les points d'arrivée des saccades correctes (Figure 2.1-c) a indiqué un effet significatif de l'EFE ($F(2,86) = 34,7$, $p < .001$, $\eta_p^2 = .45$). Les points d'arrivée étaient plus bas pour les visages joyeux ($M \pm SD$: -0,68 ± 0,54) que pour les visages apeurés ($M \pm SD$: -0,76 ± 0,54 ; $p < .001$) ou neutres ($M \pm SD$: -0,63 ± 0,52 ; $p < .001$), et pour les visages apeurés que neutres ($p = .001$). Une visualisation des points d'arrivée en fonction de l'EFE de la cible est proposée sur la Figure 2.1-d.

2.3.4 Discussion

Les résultats de cette étude ont révélé que les visages émotionnels, joyeux ou apeurés, étaient plus souvent la cible de la première saccade que les visages neutres. Ce premier résultat suggère que, même lorsqu'elles ne sont pas pertinentes, les émotions du visage peuvent faciliter la détection et le déploiement du regard vers une cible. De plus, les visages émotionnels joyeux attiraient la première saccade plus que les visages émotionnels apeurés. L'avantage des visages joyeux pourrait s'expliquer par le fait qu'ils sont plus facilement reconnaissables que les visages apeurés, et se distinguent mieux des visages neutres. Nous nous attendions à une interaction entre la cible et l'émotion, soulignant un avantage des visages émotionnels plus fort lorsque la cible était un visage féminin. Nous nous attendions également à une interaction entre le genre des participants et la cible, soulignant un avantage dans la détection de visages correspondant au genre des participants. La première interaction, entre la cible et l'émotion, n'a pas été observée. Ainsi, l'avantage des visages émotionnels n'était pas plus fort avec un visage féminin. La seconde interaction, entre le genre des participants et la cible, était seulement marginalement significative sur la proportion de saccades correctes, ce qui ne nous permet pas de conclure. Notons néanmoins que les différences de moyennes vont dans le sens de nos hypothèses, ainsi, il est possible qu'un manque de puissance explique cette absence d'effet significatif.

Bien que nous ayons observé certains effets attendus sur les proportions de saccades correctes, aucun effet n'a été observé sur les latences des saccades. Ainsi, le temps que les participants ont mis pour répondre correctement était similaire, quel que soit la cible, son expression ou le genre des participants. En moyennes, les saccades étaient déclenchées 199 ms après l'apparition des images. Ce temps de réponse est plus long que celui associé à la détection de visages (Expérience 1 de l'Article 1, autour de 176 ms), mais plus court que celui associé à la détection des émotions (Expérience 2 de l'Article 1, autour de 249 ms). Cela signifie que, même avant leur détection explicite, les émotions peuvent interférer avec le traitement du genre. L'analyse des points d'arrivée des saccades a mis en évidence des fixations plus basses sur une cible joyeuse ou apeurée que sur une cible neutre, et sur une cible joyeuse que sur une cible apeurée. Ces résultats sont en accord avec les conclusions de l'Article 1, qui soulignent l'aspect automatique de la répartition de l'attention dans un visage selon l'expression. Ainsi, même si la tâche peut orienter l'attention différemment dans le visage (par exemple plus bas dans le cadre de la détection d'émotion que dans

le cadre de la détection de visages ; la détection du genre se plaçant entre les deux), l'expression faciale va induire des décalages indépendamment de cette tâche.

Dans cette expérience, nous avons montré que les visages émotionnels peuvent faciliter la détection et le déploiement du regard vers une cible même lorsque l'expression faciale n'est pas pertinente à la tâche. Néanmoins, nous ne pouvons pas suggérer que cet effet s'applique pour toutes les tâches après l'étape de la détection des visages. Par exemple, dans des tâches dans lesquelles les visages ne sont pas pertinents, ou d'autres tâches dans lesquelles les expressions faciales ne sont pas pertinentes. Certaines études n'ont pas observé de traitement privilégié des visages émotionnels, même sur des comportements tardifs, et soulignent l'aspect conditionnel de la capture de l'attention par les visages émotionnels. Par exemple, des tâches dans lesquelles un visage émotionnel est brièvement présenté en amorce, suivi d'un stimuli à catégoriser (Koster et al., 2007 ; Puls et Rothermund, 2018 ; Victeur et al., 2019). Aussi, dans une série de quatre expériences menée par Tannert et Rothermund (2020), des participants se sont vus présenter simultanément un visage cible et un (ou plusieurs) visages distracteurs. Ils ont observé un effet des émotions seulement dans une expérience dans laquelle deux visages étaient présentés et les participants devaient indiquer si le visage cible (masculin ou féminin) était émotionnel ou neutre. Lorsque les participants devaient indiquer si le visage cible était vieux ou jeune, ou indiquer si un visage présenté en vision centrale et entouré de distracteurs était émotionnel ou neutre, cet effet n'était plus observé. Les auteurs ont suggéré que les visages émotionnels distracteurs vont capturer l'attention seulement lorsque le traitement des expressions faciales et le traitement de tous les stimuli sont rendus obligatoires. Dans notre expérience, nous avons observé un effet des émotions dans une tâche ou leur traitement n'était pas obligatoire, mais cela peut être restreint au traitement du genre, ou à l'utilisation de réponses saccadiques qui peuvent être plus adaptées pour déceler des effets émotionnels (Bannerman, Milders et Sahraie, 2009).

Chapitre 2 - Points clés

- Mise en place de trois expériences comportementales et d'une simulation.
- Les saccades vers les visages sont effectuées plus rapidement et sont plus souvent dans la bonne direction que les saccades vers les véhicules, mais cet effet n'est pas favorisé par la présence d'une expression faciale émotionnelle.
- Les saccades vers les visages émotionnels sont effectuées plus rapidement que celles vers les visages neutres. Elles sont aussi plus souvent dans la bonne direction, d'autant plus lorsque le visage émotionnel cible est joyeux plutôt qu'apeuré.
- L'avantage des visages joyeux pourrait s'expliquer par les propriétés statistiques des images, qui les rendent plus faciles à distinguer des visages neutres.
- Les saccades vers les visages féminins ou masculins sont plus souvent dans la bonne direction lorsque le visage cible est apeuré ou joyeux plutôt que neutre, et joyeux plutôt qu'apeuré.
- La configuration spatiale des expressions faciales induit des décalages automatiques de l'attention vers différentes parties du visage, qui se traduisent par des points d'arrivée des saccades plus bas pour les visages joyeux qu'apeurés, quelle que soit la tâche des participants.

Chapitre 3

Détection de visages émotionnels : influence des fréquences spatiales, du contraste, et visualisation des régions diagnostiques

Table des matières

3.1 Préface

Comme nous l'avons vu dans le chapitre précédent, particulièrement dans l'Expérience 2, les visages émotionnels sont mieux détectés que les visages neutres. Ainsi, lorsqu'un visage émotionnel est opposé à un visage neutre, les participants sont plus rapides et font moins d'erreurs lorsqu'ils doivent faire une saccade vers le visage émotionnel plutôt que vers le visage neutre. Le but de ce troisième chapitre est de préciser l'origine de cet effet. Dans un premier temps, nous nous sommes interrogés sur le rôle des fréquences spatiales dans la détection de visages émotionnels ou neutres. En effet, les modèles de traitement des émotions suggèrent que les stimuli liés à une menace, tels que les visages apeurés, peuvent être détectés sur la base d'une extraction rapide des BFS (par exemple LeDoux, 1998 ; Tamietto et de Gelder, 2010). Cependant, ce point de vue reste débattu. Par exemple, selon d'autres auteurs le décodage des expressions faciales se produit avec une utilisation plus flexible des fréquences spatiales (par exemple Morrison et Schyns, 2001 ; Schyns et Oliva, 1999). De plus, les expériences que nous avons présentées dans le chapitre précédent, avec des images non filtrées, n'ont pas mis en évidence un traitement plus rapide des visages apeurés. Dans ce chapitre, nous avons reproduit l'Expérience 2 du chapitre précédent (Visage émotionnel versus Visage neutre), en modifiant le contenu fréquentiel de nos images de manière à conserver uniquement les HFS ou les BFS. Afin de prendre en compte les différences de contraste entre les HFS et les BFS (voir section 1.8.3 du Chapitre 1), nous avons comparé les résultats de deux groupes de participants : un groupe pour lequel le contraste n'était pas égalisé et un groupe pour lequel il était égalisé après le filtrage. Dans un second temps, nous nous sommes interrogés sur le rôle des différentes parties du visage (en particulier les yeux et la bouche) dans la détection

de visages émotionnels ou neutres. Pour cela, nous avons d'abord utilisé un modèle de saillance *bottom-up*, afin de quantifier la saillance de chaque pixel à partir de facteurs physiques. Plus particulièrement, nous avons testé la corrélation entre la saillance moyenne obtenue dans la région des yeux ou de la bouche et les performances des participants dans l'expérience présentée dans ce chapitre. Ensuite, nous avons utilisé un réseau de neurones convolutionnel (ou *convolutional neural network* ; CNN), entraîné à une tâche similaire à celle des participants, pour mettre en avant les régions utilisées par le réseau pour faire la tâche. Nous avons ainsi obtenu des cartes de saillance spécifiques à une tâche, contrairement aux premières cartes générées par les modèles de saillance *bottom-up* qui sont elles indépendantes de la tâche. À partir de ces deux types de cartes (les cartes *bottom-up* et les cartes basées sur le CNN) nous avons quantifié la capacité de la saillance de la bouche à prédire les performances des participants dans la tâche comportementale.

Dans ce chapitre, nous présenterons d'abord les résultats obtenus pour l'expérience comportementale et l'analyse de la saillance *bottom-up*, qui ont fait l'objet d'un article soumis (Article 2). Ensuite, nous présenterons le modèle CNN qui a permis de générer des cartes de saillance spécifique à la tâche, qui fait l'objet d'un autre article soumis (Article 3). Les annexes de ces articles sont présentées à la fin du manuscrit de thèse, dans l'Appendice B.

3.2 Article 2

Dans l'étude présentée dans ce second article, nous nous sommes intéressés au rôle des fréquences spatiales dans la détection et le déclenchement de saccades vers des visages émotionnels ou neutres. Cet article est composé d'une expérience comportementale en choix saccadique et d'une analyse de la saillance *bottom-up* de nos stimuli. Dans l'expérience comportementale, nous avons utilisé un paradigme de choix saccadique opposant un visage émotionnel (joyeux ou apeuré) et un visage neutre, conformément à ce qui a été fait dans l'Expérience 2 de l'Article 1. Nous avons cette fois manipulé le contenu fréquentiel des images, qui pouvaient être présentées en HFS, en BFS ou non filtrées. Les participants devaient effectuer une saccade le plus rapidement possible vers le visage émotionnel dans une session ou vers le visage neutre dans une autre session. Un premier groupe de participants a été testé sur des stimuli pour lesquels le contraste des images n'était pas égalisé, tandis qu'un second groupe a été testé sur des stimuli pour lesquels le contraste était égalisé.

Nous nous attendions dans un premier temps à répliquer les résultats de l'Expérience 2 de l'Article 1 avec des images non filtrées. Plus spécifiquement, nous nous attendions à des proportions de saccades correctes plus élevées et des latences plus courtes lorsque la cible est le visage émotionnel plutôt que le visage neutre. Lorsque la cible est le visage émotionnel, nous nous attendions à des proportions de saccades correctes plus élevées lorsqu'il était joyeux plutôt qu'apeuré. Concernant l'effet des fréquences spatiales, nous supposions que, si les BFS sont suffisantes pour discriminer les expressions faciales et sont traitées plus rapidement que les HFS, les latences devraient être plus courtes pour

les stimuli en BFS qu'en HFS. Nous nous attendions également, lorsque la cible était un visage émotionnel, à une interaction entre les fréquences spatiales et l'expression faciale. Plus précisément, nous nous attendions à observer de meilleures performances pour les visages joyeux qu'apeurés avec des images non filtrées. Mais, pour les images en BFS, en supposant que les BFS peuvent déclencher une réponse cérébrale rapide qui facilite la détection des visages apeurés, nous nous attendions à l'inverse. Nous nous attendions aussi à ce que les performances soient modulées par la condition de contraste, sous la forme d'une interaction entre les fréquences spatiales et le groupe de contraste. Plus précisément, nous nous attendions à ce que les performances en BFS soient meilleures en condition de contraste non égalisé ; étant donné que les images en BFS non égalisées présentent un contraste plus élevé que les images HFS non égalisées. Lorsque le contraste est égalisé, nous supposions que l'avantage attendu des BFS sur les latences des saccades serait diminué.

Les résultats de cette expérience ont montré que, conformément aux résultats observés dans l'Article 1, les performances étaient meilleures (c'est-à-dire que les proportions de saccades correctes étaient plus élevées et les latences plus courtes) lorsque la cible était un visage émotionnel plutôt qu'un visage neutre. Pour les cibles émotionnelles, la proportion de saccades correctes était plus élevée lorsque le visage était joyeux plutôt qu'apeuré, un effet principalement observé quand les images étaient en HFS. De manière globale, la proportion de saccades correctes était plus élevée pour les images en HFS et non filtrées que pour les images en BFS. Et, l'avantage des HFS sur les BFS était principalement observé avec un visage joyeux. En ce qui concerne les latences des saccades, elles étaient globalement plus courtes pour les images non filtrées que pour les images filtrées, mais aussi pour les images en HFS que pour les images en BFS. Enfin, l'impact de l'égalisation du contraste sur les performances n'a pas été aussi important que prévu. Cependant, une interaction marginale entre les fréquences spatiales et le contraste observée sur les latences des saccades a montré que la différence entre les HFS et les BFS était significative seulement pour les images avec un contraste égalisé. Ce résultat suggère que le traitement des fréquences spatiales peut être dépendant des différences de contraste.

L'analyse de saillance nous a permis de mettre en avant les zones de nos stimuli qui se distinguent particulièrement, sur la base des statistiques de luminance des pixels. Conformément à ce qui a été fait dans une autre étude (Calvo et Nummenmaa, 2011), le but de cette analyse était de quantifier le lien entre la saillance de différentes parties du visage et la performance des participants. L'hypothèse sous-jacente et que, si une partie du visage est utile à la tâche (ici, la discrimination d'un visage émotionnel et d'un visage neutre), plus elle est saillante dans une condition (par exemple en HFS avec un visage joyeux), plus les performances seront bonnes dans cette condition. Les yeux et la bouche étant des régions connues pour être importantes dans le décodage des émotions, nous nous attentions à ce que la saillance des yeux et de la bouche dans une condition soit liée à la performance des participants dans cette condition. Les résultats de cette analyse ont montré que la saillance de la bouche corrèle significativement avec les résultats comportementaux. Ainsi, plus la bouche est saillante, en moyenne, dans une condition expérimentale (c'est-à-dire, une condition de cible, de fréquences spatiales, d'émotion et

de contraste), plus les performances seront bonnes dans cette condition. Au contraire, la saillance des yeux ne semble pas très importante pour cette tâche. En résumé, les résultats de cette expérience n'ont pas mis en avant un traitement privilégié des visages apeurés, en BFS. En lien avec l'idée d'une utilisation flexible des fréquences spatiales, nous suggérons que dans cette tâche l'information transmise par les HFS est plus pertinente. De plus, ces résultats supposent que les performances des participants pour détecter et faire une saccade vers un visage émotionnel ou neutre peuvent être expliquées, au moins en partie, par les statistiques des images au niveau de la région de la bouche.

L'Article 2 a été soumis dans la revue *Brain and Cognition* en décembre 2021. Le Tableau 3.1 dresse la liste des contributions de chaque auteur.

Contributeurs	Contributions
Léa Entzmann	Conception de l'expérience ; Recueil des données ; Analyse des données ; Rédaction du manuscrit ; Édition du manuscrit
Nathalie Guyader	Conception de l'expérience ; Édition du manuscrit
Louise Kauffmann	Conception de l'expérience ; Édition du manuscrit
Carole Peyrin	Conception de l'expérience ; Édition du manuscrit
Martial Mermillod	Conception de l'expérience ; Édition du manuscrit

Table 3.1 – Contributions des auteurs de l'Article 2.

Detection of emotional faces: the role of spatial frequencies and local features

Léa Entzmann[1, 2] Nathalie Guyader[2], Louise Kauffmann[1], Carole Peyrin[1], & Martial Mermillod[1]

[1] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000, Grenoble, France.

[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France.

Corresponding author
Lea Entzmann
LPNC
1251 Avenue Centrale,
Batiment Michel Dubois,
38400 Saint-Martin-d'Heres
lea.entzmann@univ.grenoble-alpes.f

Abstract

Models of emotion processing suggest that threat-related stimuli such as fearful faces can be detected based on the rapid extraction of low spatial frequencies. However, this remains debated as other frameworks argue that the decoding of facial expressions occurs with a more flexible use of spatial frequencies. The purpose of this study was to clarify the role of spatial frequencies and differences in luminance contrast between spatial frequencies, on the detection of facial emotions. We used a saccadic choice task in which emotional-neutral face pairs were presented and participants were asked to make a saccade toward the neutral or the emotional (happy or fearful) face. Faces were displayed either in low, high, or broad spatial frequencies. Results showed that participants were better to saccade toward the emotional face. They were also better for high or broad than low spatial frequencies, and the accuracy was higher with a happy target. An analysis of the eye and mouth saliency of our stimuli revealed that the mouth saliency of the target correlates with participants' performance. Overall, this study underlines the importance of local more than global information, and of the saliency of the mouth region in the detection of emotional and neutral faces.

Keywords: Emotional Facial Expressions; Eye Movements; Spatial Frequencies; Contrast; Visual Saliency

1. Introduction

In everyday life, the ability to detect a face and more specifically the emotion of a face, is crucial, as emotional faces convey essential information for appropriate social behaviour. Thereby, the human visual system developed specific mechanisms to efficiently process faces (Haxby et al., 2000; Liu et al., 2002; Zhao et al., 2018). Several behavioural experiments highlighted the preferential processing of faces stimuli (Cerf et al., 2009; Coutrot & Guyader, 2014; Farah et al., 1998; Langton et al., 2008). For example, some studies using eye-tracking showed that faces can be detected as early as 100 ms, whereas more time is needed for the detection of other stimuli, such as vehicles or animals (Crouzet, 2010; Entzmann et al., 2021; Kauffmann et al., 2019, 2021).

This efficient and fast processing of face stimuli might be explained by the fact that face detection relies on coarse more than fine information (Awasthi et al., 2011; Goffaux et al., 2011; Goffaux & Rossion, 2006; Guyader et al., 2017; Peters et al., 2018; Quek et al., 2018). In fact, visual perception is thought to be based on the parallel extraction of different visual features on different spatial frequencies and to follow a default, predominantly coarse-to-fine processing sequence (Bar, 2003; Hegde, 2008; Kauffmann et al., 2014; Kauffmann, Chauvin, et al., 2015; Musel et al., 2012; Petras et al., 2019; Peyrin et al., 2010; Schyns & Oliva, 1994). Low spatial frequencies (LSF; carrying coarse information) would be extracted first and rapidly processed allowing to form a coarse representation of the visual input, while high spatial frequencies (HSF), which convey finer information (e.g., details and edges), would be carried more slowly and used to refine the first LSF-based analysis. Several studies highlighted the primary role of LSF information in face processing (Awasthi et al., 2011; Goffaux et al., 2011; Goffaux & Rossion, 2006; Guyader et al., 2017; Peters et al., 2018; Quek et al., 2018). For example, Goffaux and Rossion (2006) showed that the whole-part advantage (i.e., superior recognition of the eyes when it is presented in the context of a whole face rather than isolated) in an identity matching task was larger with LSF and broad spatial frequency (BSF; i.e., unfiltered images) than in HSF images. Moreover, Guyader et al. (2017) presented simultaneously a face and a vehicle image, both presented without filtering, in HSF, or in LSF, and asked participants to make a saccade toward the face. They observed that within 130-140 ms participants were able to make more correct than incorrect saccades to faces without filtering or presented in LSF, whereas they required more time for images presented in HFS.

Whereas face processing mainly relies on the extraction of LSF information, the role of spatial frequencies in the detection of facial expressions is less clear. Facial expressions provide indications on the emotional state of others, and consequently, on the identification of potential threats. Classical views of the visual processing of emotional stimuli in the brain suggest that LSF

play a crucial role, especially for the perception of threat-related stimuli, such as fearful faces. This hypothesis relies on the existence of a short and direct superior colliculus-pulvinar pathway to the amygdala which enable rapid detection of threat (LeDoux, 2000; Méndez-Bértolo et al., 2016; Morris, 1998; Öhman, 2005; Tamietto & de Gelder, 2010; Vuilleumier et al., 2003). This subcortical pathway is supposed to transmit LSF information in parallel to a slower cortical pathway that transmits finer information. This view is notably supported by an fMRI study, which showed higher activation in the amygdala for fearful faces presented in LSF than in HSF, and an activation of the pulvinar and superior colliculus by fearful faces only when they were presented in LSF (Vuilleumier et al., 2003). Using intracranial EEG recordings in the amygdala, fearful faces were also found to evoke early activity, 75 ms post-stimulus onset, only when they were unfiltered or in LSF (Méndez-Bértolo et al., 2016). Overall, these neuroimaging studies suggest that facial expressions can be detected rapidly, through LSF information only. In this sense, computational studies showed that the information carried by LSF is sufficient to discriminate facial expressions (Mermillod et al., 2009, 2010). For instance, Mermillod et al. (2010) tested the usefulness of LSF compared to HSF and BSF information, in a facial expression categorization task performed by an artificial neural network. They found that LSF images lead to better categorization of facial expressions (anger, disgust, fear, sadness, happiness and surprise).

However, the existence of a subcortical pathway for the rapid detection of fearful faces, as well as the crucial role of LSF, is still a matter of debate (Pessoa & Adolphs, 2010). For example, McFadyen et al. (2017) used magnetoencephalography to measure neural activity while participants performed a gender discrimination task of neutral and fearful faces in LSF or HSF. They demonstrated through dynamic causal modeling that the most likely underlying subcortical neural network consisted of a pulvinar-amygdala connection that was neither influenced by spatial frequencies nor by emotions, in line with other results (Fitzgerald et al., 2006; Garvert et al., 2014; Johnson, 2005). Moreover, several studies argued for a more flexible use of spatial frequencies to decode facial expressions, occurring later in the time course of visual processing (Schyns et al., 2009). Different spatial frequency bands would be used depending on the facial expression (Morrison & Schyns, 2001; Oliva & Schyns, 1997; Smith & Schyns, 2009) and the task (Schyns & Oliva, 1999; Smith & Merlusca, 2014). For example, fearful face categorization might rely mostly on the wide-opened eyes and therefore on the extraction of HSF, whereas a larger scale would be used for other expressions (Adolphs et al., 2005; M. L. Smith et al., 2005; Stein et al., 2014). However, such behavioural studies on the role of spatial frequencies in the processing of facial expressions used manual responses and may reflect processes occurring at a later stage of visual processing than what is described in neuroimaging studies for the fast fear detection.

In a previous study, we tested the detection of emotional and neutral faces using a saccadic

choice task. More precisely, emotional-neutral pairs of faces were presented to participants, who were asked to perform a saccade as fast and as accurately as possible toward the emotional (happy or fearful) or the neutral face. The use of saccadic eye movements as a behavioural response was motivated by the fact that the latency of saccadic response has been shown to provide more precise and robust measures of visual processing speed than what allows manual responses (Kirchner & Thorpe, 2006). We found that participants were faster and made less errors when they had to saccade toward the emotional than the neutral face. This was especially the case when the emotional face was happy. Using an artificial neural network, we showed that this advantage for happy faces can, at least partly, be explained by perceptual factors. More precisely, the network was trained and tested on its ability to discriminate a neutral from an emotional face, and performed better when the emotional face was happy rather than fearful. We also analyzed the saccade endpoints and we observed that saccades landed lower on happy than fearful faces, suggesting that local features like the mouth or the eyes attracted the gaze differently according to the emotional expression. Indeed, the eyes and the mouth are the most useful parts to discriminate facial expressions (Eisenbarth & Alpers, 2011; Smith & Schyns, 2009; Wegrzyn et al., 2017). It is likely that the saliency of these regions, or their respective diagnosticity, varies according to the expression, and shifts the endpoints of the saccades. In the present paper, we reproduced the eye-tracking experiment introduced in Entzmann et al. (2021), but with different spatial frequency conditions. This can be seen as an equivalent to what Guyader et al. (2017) did for face detection, as they use a saccadic choice task with face-vehicle pairs presented in HSF, LSF, and BSF.

Our purpose was first to clarify the role of spatial frequencies on the detection of emotional faces. We also considered methodological issues about the filtering procedure that may explain discrepancies between studies (Kauffmann, Chauvin, et al., 2015; Perfetto et al., 2020; Vlamings et al., 2009). Filtered images, like LSF and HSF images, differ not only in their spatial frequency content but also in their luminance contrast (i.e., the magnitude of luminance variation in a stimulus relative to its mean luminance, more simply referred to as contrast; Shapley & Enroth-Cugell, 1984). This is due to the fact that the luminance contrast in scenes decreases as spatial frequency increases (Field, 1987), leading LSF images to have higher luminance contrast than HSF. Such luminance contrast differences between HSF and LSF can influence the processing of the spatial frequency content in visual stimuli. Indeed, previous studies have shown that high contrast stimuli can be detected faster than low-contrast stimuli (e.g., Ludwig et al., 2004). Actually, in studies using visual scenes as stimuli (Kauffmann, Chauvin, et al., 2015; Kauffmann, Ramanoël, et al., 2015), the temporal advantage of LSF on HSF processing, as well as the predominant coarse-to-fine processing of spatial frequencies, was attenuated by the equalization of luminance contrast between LSF and HSF. This suggests that the higher luminance contrast of LSF may contribute to the

classical effect of the predominance of LSF on HSF information. Importantly, such methodological issue may also explain discrepancies observed in past studies investigating the role of spatial frequencies for emotional face processing. For example, McFadyen et al. (2017) used LSF and HSF faces that were equalized in luminance contrast, unlike previous neuroimaging studies (e.g., Méndez-Bértolo et al., 2016; Vuilleumier et al., 2003). This may explain the absence of selectivity of the amygdala to LSF fearful faces and suggests that higher response in the amygdala for LSF faces may in fact be due to their higher contrast.

We used a saccadic choice task with emotional-neutral face pairs, in which we manipulated the spatial frequency content (LSF, HSF or BSF) of images. Participants were asked to make a saccade as rapidly as possible toward the emotional (happy or fearful) or the neutral face. A first group of participants was tested on stimuli (LSF, HSF and BSF) for which the luminance contrast across faces was not equalized (NonEQ condition) whereas a second group was tested on stimuli (LSF, HSF and BSF) for which this luminance contrast was equalized (using a root-mean-square contrast normalization; EQ condition). We chose this paradigm, where two faces are opposed to each other, because it allowed us to show in our previous study that, for unfiltered images, emotional faces are better detected than neutral faces (Entzmann et al., 2021). This effect is not systematically found with other paradigms (Devue & Grimshaw; 2017). For example, if we oppose a face and a vehicle and ask participants to make a saccade towards the face, they perform equally whether the face is emotional or neutral (Entzmann et al., 2021).

We expected better performances (higher accuracy, shorter latencies) when participants were asked to saccade toward the emotional compared to the neutral face. In addition, we supposed that, if LSF are sufficient to discriminate facial expressions and are processed faster than HSF, latencies should be shorter for LSF than HSF stimuli. We also expected, when the target was an emotional face, an interaction between the spatial frequency content and the emotional facial expression. More specifically, we expected better performance for happy than fearful face targets for BSF images, as previous studies using unfiltered images found an advantage in the detection of happy faces (Calvo & Nummenmaa, 2009, 2011; Entzmann et al., 2021). However, for LSF images, we expected the opposite (i.e., an advantage in the detection of fearful faces), as LSF may trigger a fast brain response that would facilitate the detection of fearful faces. We also expected performance to be modulated according to the contrast condition (NonEQ vs. EQ group), in the form of an interaction between the spatial frequency content and the contrast group. More precisely, we expected that if the fast detection of LSF fearful faces is also explained by their higher contrast, the better detection of LSF over HSF faces should be enhanced in the NonEQ group. However, for the EQ group the difference between LSF and HSF faces should be attenuated, or could even be inverted toward an advantage of HSF. We also expected, when the target was an emotional face, an

interaction between the spatial frequency content, the contrast group and the emotional facial expression, as we supposed that the better detection of fearful than happy faces in LSF would be reduced or disappear in the EQ group.

To see if local statistical differences in the saliency of some diagnostic parts of the face, namely the mouth or the eyes, could explain participants' behaviour, we also computed the saliency maps of our face stimuli (Borji & Itti, 2013; Foulsham et al., 2008; Itti & Koch, 2000; Marat et al., 2009). Indeed, we supposed that the conditions in which the mouth or the eyes of the target were the most salient, could be the conditions in which the task was the easiest. Visual saliency can be viewed as the intensity with which a region will attract attention, independently of task demands. Several studies showed that the mouth and the eyes play a critical role in the decoding of facial expressions (Eisenbarth & Alpers, 2011; Smith & Schyns, 2009; Wegrzyn et al., 2017), in happy face detection (Calvo & Nummenmaa, 2011) or emotion detection (Sweeny et al., 2013). For instance, Calvo and Nummenmaa (2011) studied the role of perceptual (e.g., luminance, mouth or eye saliency differences), as well as higher level factors (e.g., valence) on the detection of happy faces in a saccadic choice task, and only found a contribution of the mouth saliency. Although Calvo and Nummenmaa found no contribution of the eye saliency in their happy face detection task, we however analyzed it in the present study, as it may still contribute to the detection of emotional or neutral faces.

2. Saccadic choice task

2.1. Materials and method

2.1.1. Participants

Eighty-one participants recruited from Grenoble Alpes University took part in the experiment. Based on our previous study (Entzmann et al., 2021), we were able to estimate a sample size that is sufficient enough to find the effect of facial emotion on saccadic responses. Using the G*Power software (Faul et al., 2007) with a power of .95 at the standard .05 alpha error probability for the main effect of the target on the accuracy, the estimate sample size was 16. However, in the present study we introduced new variables, with manipulation on the spatial frequency content and contrast of our stimuli. We did not find any study that allows a precise estimate of the strength of the expected effects (i.e., a study with facial expressions, saccadic responses, and spatial frequencies or contrast). Therefore, we choose to consider a larger sample size (approximately 40 participants per group) to increases the statistical power. Three participants were removed from statistical analysis due to a low proportion of correct responses (below 50% in

each session), or a high proportion of invalid trial (above 50%), leading to a group of seventy-eight participants included in our data analysis (39 females; mean age ± SD: 21.39 ± 0.98 years; age range: 18–33 years). They all had normal or corrected-to-normal vision and gave their informed written consent before the experiment, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.1.2. Stimuli

Stimuli consisted of 60 grayscale face photographs portraying 20 different individuals (10 women) with 3 emotions (happy, fearful or neutral); images had a resolution of 300 × 300 pixels and were chosen among the Karolinska Directed Emotional Faces database (Lundqvist et al., 1998). The mean luminance (i.e., mean pixel intensity) as well as the root mean squared contrast (RMS contrast; corresponding to the standard deviation of pixel intensity) were equalized among all the images prior to any further manipulation to obtain a mean luminance of 125 and a mean RMS contrast of 46 (for pixel intensity values comprised between [0,255]; such values corresponded to the mean luminance and contrast values of all the stimuli). Each image was then filtered with a low- or a high-pass filter with cut-off frequencies respectively set to 1 cycles/degree (11 cycles/image for LSF) and 6 cycles/degree (66 cycles/image for HSF). These values were chosen to match a previous study (Guyader et al., 2017) that measured the influence of spatial frequencies on fast saccades toward faces using a saccadic choice task. Therefore, each image was viewed under three different spatial frequency conditions: a HSF condition, a LSF condition and a BSF condition. Because the amplitude spectrum of natural images decreases as spatial frequency increases (Field, 1987; Kauffmann, Chauvin, et al., 2015; van der Schaaf & van Hateren, 1996), LSF images have a higher RMS contrast than HSF images. Hence, to dissociate the respective contributions of spatial frequency and luminance contrast, two sets of stimuli were built. In one stimulus dataset, a RMS contrast equalization was applied after the filtering process (EQ condition), whereas in the other stimulus dataset there was no contrast equalization after filtering (NonEQ condition). In the EQ dataset, all images, filtered and unfiltered, had the same mean luminance contrast (set to the value of 46, which corresponds to the contrast of BSF images), whereas in the NonEQ dataset the mean luminance contrast of HSF, LSF, and BSF images was left unchanged (mean RMS contrast for HSF images: 8; mean RMS contrast for LSF images: 69; mean RMS contrast for BSF images: 46; all values comprised between [0,255]; see Fig. 1a for examples of stimuli).

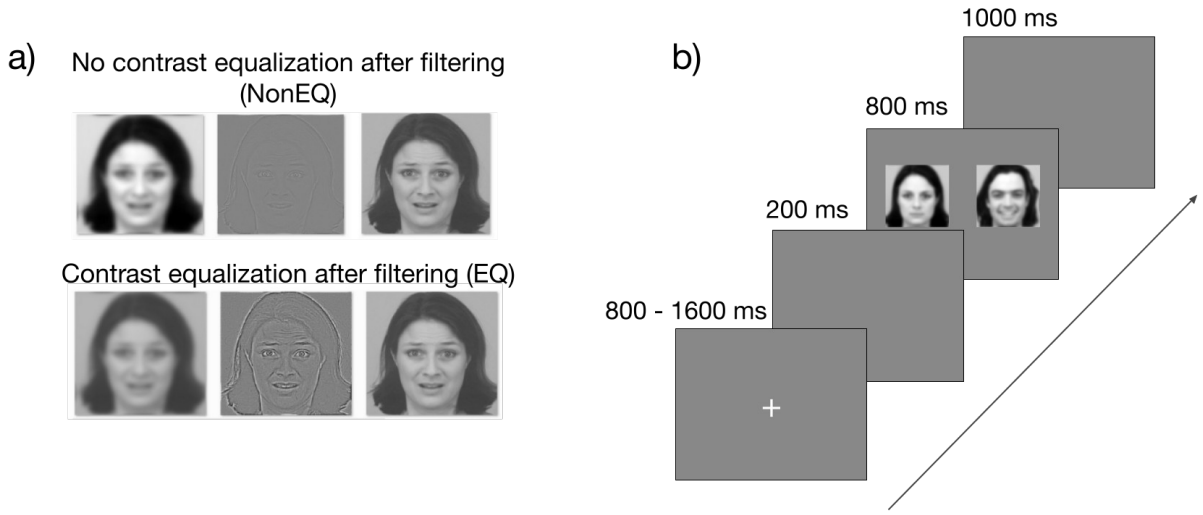

Figure 1. (a) Example of an image in the different contrast and spatial frequency conditions: In the first raw, the non-equalized (NonEQ) contrast condition with LSF, HSF, and BSF images. In the second raw, the contrast equalized (EQ) condition with LSF, HSF, and BSF images. In the EQ condition, all images have a mean RMS contrast of 46, for pixel intensity values comprised between [0,255]. In the NonEQ condition LSF, HSF and BSF images have a mean RMS contrast of 68, 8, and 46, respectively. Mean luminance was set to 125 and was the same in all conditions. (b) Time course of a trial: A central fixation cross is displayed during 800 to 1600 ms, followed by a 20-ms gap. Then two images, an emotional and a neutral face are displayed for 800 ms, followed by a 1000 ms inter-stimulus interval.

2.1.3. Materials

Stimuli were displayed on a 24-inch screen with a spatial resolution of 1360×768 pixels and a refresh rate of 60 Hz. Eye movements were recorded with an Eyelink 1000 Plus (SR Research) eye-tracker with a 1000 Hz sampling frequency. Viewing was binocular, but only the position of the dominant eye was recorded (the left eye for 19 participants, and the right eye for 59 participants). Saccades were automatically detected by the Eyelink software based on a minimum velocity of 30 degrees/s, a minimum acceleration of 8000 degrees/s2 and a minimum motion of 0.15 degrees. Blinks were detected when the pupil was partially or totally occluded, and fixations were detected when there was no blink and no saccade in progress.

2.1.4. Procedure

The experiment was divided into two sessions of 240 trials each, whose order was counterbalanced between participants. In one session, the target stimulus was the emotional face

(happy or fearful; the distractor was the neutral face) while in the other session, the target stimulus was the neutral face (the distractor was the emotional face, which was either happy or fearful). Each session was divided into blocks of 80 trials each, corresponding to the different spatial frequency conditions (HSF, LSF, and BSF) presented in a randomized order. One group of participants (n=37) performed the task with contrast equalization (EQ condition) of stimuli whereas another group (n=41) performed the task without contrast equalization (NonEQ condition). A calibration was performed at the beginning of each session and in each session, at the beginning of each block (every 80 trials). A drift correction was applied every ten trials (if the drift was larger than 1° a new calibration was done). During the calibration, participants were asked to gaze at 9 white dots appearing sequentially in a 3 × 3 grid covering the entire screen. Matlab (MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997) were used to control timing and stimulus display, as well as communication with the eye-tracker.

During the experiment, participants were seated in a semi-lighted room, with their head stabilized by a forehead-rest and a chin-rest at a fixed distance of 57 cm away from the screen in order to respect a stimulus size of 11×11 degree of visual angle. A session lasted approximately 20 minutes and the whole experimental procedure took approximately 50 minutes. For each trial, participants were asked to fixate a white cross presented during a pseudo-random time interval ranging between 800 and 1600 ms. After a 200 ms gap, two images (an emotional and a neutral face) were simultaneously displayed (one on each side of the screen) during 800 ms. Participants were asked to make a saccade as fast as possible toward the target image. The target image was given to participants at the beginning of each session: the emotional or the neutral face. The center of each image was located at a fixed distance of 8° of eccentricity from the center of the screen. A trial ended with the presentation of a gray background during 1000 ms (Fig. 1b). Twelve practice trials (including different images than those used in the experiment) were set up in a training session at the beginning of the experiment to allow participants to be familiar with the task.

2.1.5. Data analysis

Preprocessing: A preprocessing was applied to eye movement data in order to eliminate non-valid trials. Valid trials were selected according to the following criteria. First, a saccade should be the first event after stimulus onset (i.e., there is no blink before the response). Second, this first saccade should have a latency greater than 50 ms (to remove anticipatory saccades), a starting point within a radius of 2° around the center of the screen, and a duration smaller than 100 ms. Moreover, the amplitude of the first saccade should be greater than 1° and should not go beyond the screen. This preprocessing led to a rejection of 13% of the initial number of trials. The

first saccade was considered as "correct" if it was directed toward the side of the display containing the target and as "error" if it was in the opposite direction (i.e. directed toward the distractor).

Statistical analyses: Statistical analyses were carried out using the open-source software R with R Studio (Racine, 2012). Mean accuracy (in proportion of correct saccades) and mean latency of correct saccadic responses (in ms) were computed for each participant in each experimental condition, and were analysed as dependent variables. First, we studied accuracy and latency differences according to the target, the spatial frequency and the contrast (i.e., we tested main effects and interactions between those factors). A mixed analysis of variance (ANOVA) was used, with the Target (Emotional, Neutral) and the Spatial Frequency (BSF, HSF, LSF) as within-subject factors, and the Contrast (EQ, NonEQ) as a between-subject factor.

Then, two other analyses were carried out, to test for the influence of the Emotional facial expression (EFE) of the target in one analysis, or the influence of the EFE of the distractor in another analysis. The first analysis was applied on trials in which the target was an emotionnal face (and the distractor was a neutral face). This analysis was set up to study the influence of the EFE of emotional face targets (i.e., to test for main effect of the EFE, and interactions with spatial frequency and contrast). A mixed ANOVA was used on trials for which the target was the emotional face. The EFE of the target (Happy, Fearful) and the Spatial Frequency (BSF, HSF, LSF) were defined as within-subject factors, and the Contrast (EQ, NonEQ) as a between-subject factor. The second analysis was then applied on trials in which the target was a neutral face (and the distractor was an emotional face). In this case, the EFE of the distractor (Happy, Fearful) and the Spatial Frequency (BSF, HSF, LSF) were defined as within-subject factors, and the Contrast (EQ, NonEQ) as a between-subject factor. An effect was considered significant if its p-value was below the threshold $\alpha = .05$, and effect sizes were estimated by calculating partial eta-squared (η_p^2). T-tests with Bonferroni corrections were used for pairwise comparisons.

2.2. Results

2.2.1. Accuracy

Mixed ANOVA performed on mean accuracy (Fig. 2a) indicated a significant effect of the Target ($F(1,76) = 67.1$, $p < .001$, $\eta_p^2 = .47$), and the Spatial Frequency ($F(2,152) = 3.89$, $p = .023$, $\eta_p^2 = .05$). The accuracy was higher when the target was emotional ($M \pm SD$: .65 ± .12) than neutral ($M \pm SD$: .59 ± .12). Also, it was higher for both BSF ($M \pm SD$: .62 ± .12; $p = .003$) and HSF ($M \pm SD$: .62 ± .12; $p = 0.01$) than LSF ($M \pm SD$: .61 ± .12) images. A marginally significant interaction between Spatial Frequency and Target ($F(2,152) = 3.1$, $p = .05$, $\eta_p^2 = .04$) was observed. The main effect of Contrast, as well as other interactions, were not significant.

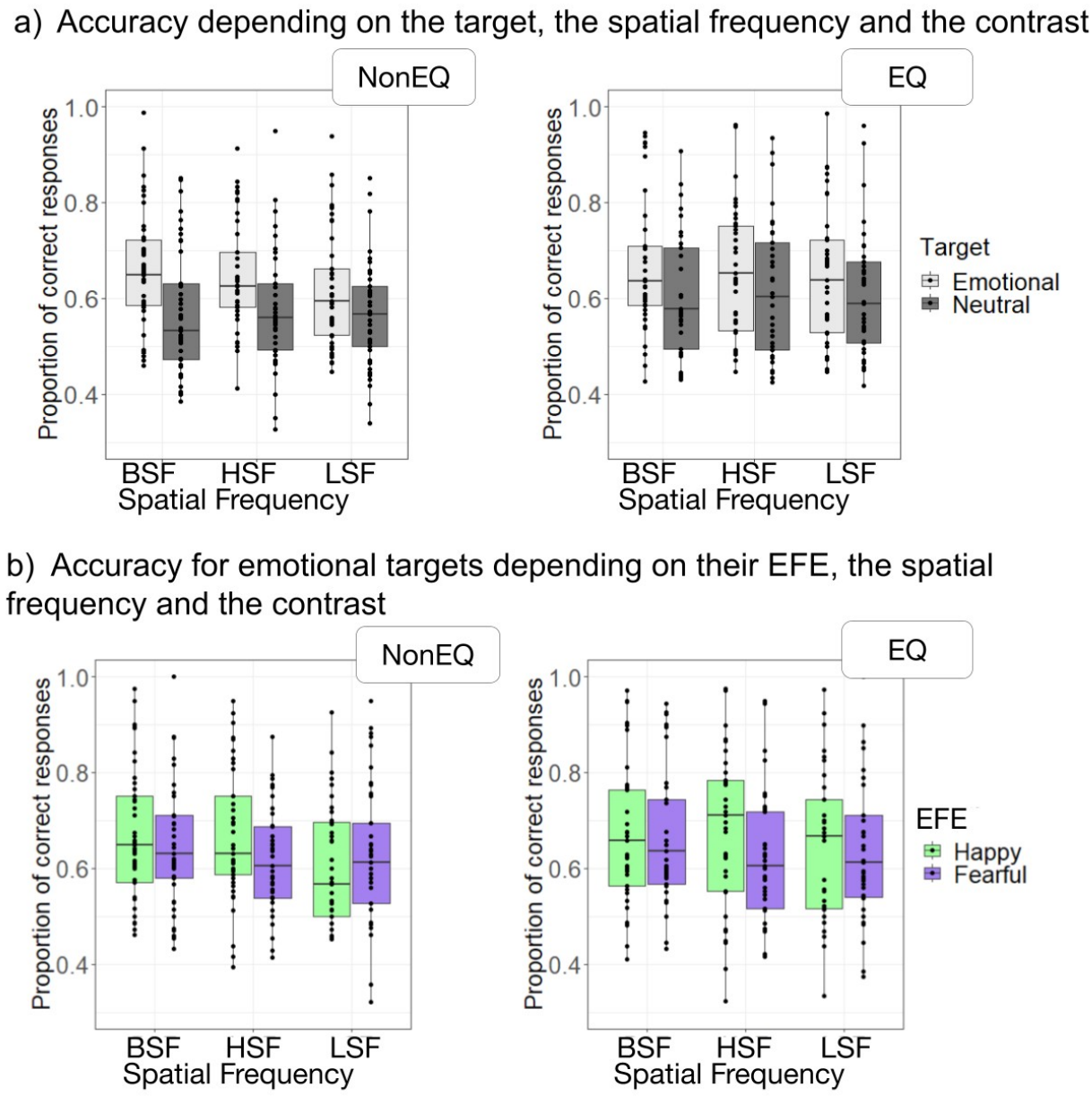

Figure 2. Boxplots for mean accuracy (proportion of correct responses) (a) for emotional (black) and neutral (gray) targets, for the three spatial frequency conditions (BSF, HSF, and LSF) and the two contrast conditions (NonEQ and EQ) and (b) for emotional targets, with a fearful (purple) or happy (green) face, for the three spatial frequency conditions (BSF, HSF, and LSF) and the two contrast conditions (NonEQ and EQ).

For the interaction between Spatial Frequency and Target, pairwise comparisons showed that when the target was neutral there was no significant effect of spatial frequencies, whereas there was a difference between BSF ($M \pm SD$: .66 ± .13) and LSF ($M \pm SD$: .60 ± .13) images when the target was emotional ($p_{corrected}$ = .02), and no significant differences between HSF and LSF or BSF images.

2.2.2. Latency

Mixed ANOVA performed on mean latency (Fig. 3a) of correct saccadic responses indicated a significant effect of the Target ($F(1,76) = 35.8$, $p < .001$, $\eta_p^2 = .31$) and the Spatial Frequency ($F(2,152) = 18.9$, $p < .001$, $\eta_p^2 = .2$). Saccades were elicited faster when the target was emotional ($M \pm SD$: 227 ± 69 ms) than neutral ($M \pm SD$: 253 ± 89 ms). Latency was also shorter for BSF ($M \pm SD$: 232 ± 74 ms) than HSF ($M \pm SD$: 238 ± 76 ms; $p = .01$) or LSF ($M \pm SD$: 250 ± 85 ms; $p < .001$) images, and for HSF than LSF images ($p < .001$). The ANOVA also revealed a significant interaction between Target and Contrast ($F(1,76)=8.2$, $p = .005$, $\eta_p^2 = .1$), and a marginally significant interaction between Spatial Frequency and Contrast ($F(1,152)=2.6$, $p = .073$, $\eta_p^2 = .034$). The main effect of Contrast, as well as other interactions, were not significant.

For the Target and Contrast interaction, pairwise comparisons showed that latencies were significantly shorter for emotional compared to neutral targets only in the EQ contrast condition (*M ± SD for emotional targets*: 229.5 ± 67.5 ms; *M ± SD for neutral targets*: 269.9 ± 96 ms; $p_{corrected} < .001$), as there was no significant difference between emotional and neutral targets in the NonEQ contrast condition. Also, for the Spatial Frequency and Contrast interaction, pairwise comparisons showed that the difference between LSF and HSF images was only significant in the EQ contrast condition (*M ± SD for HSF*: 244.1 ± 81.1 ms; *M ± SD for LSF*: 264 ± 87.3 ms; $p_{corrected} < .001$).

2.2.3. Analysis as a function of the EFE of emotional face targets

Accuracy: For emotional face targets, mixed ANOVA performed on mean accuracy (Fig. 2b) indicated a significant main effect of the EFE ($F(1,76) = 8.4$, $p = .004$, $\eta_p^2 = .1$), and the Spatial Frequency ($F(2,152) = 5.8$, $p = .003$, $\eta_p^2 = .07$). Performances were better when the target was happy ($M \pm SD$: .66 ± .13) than fearful ($M \pm SD$: .64 ± .12). They were also better for both BSF ($M \pm SD$: .66 ± .13; $p < .001$) and HSF ($M \pm SD$: .65 ± .13; $p = 0.036$) than LSF ($M \pm SD$: .63 ± .13) images. A significant interaction between Spatial Frequency and EFE ($F(2,152) = 7.2$, $p = .001$, $\eta_p^2 = .087$) was observed. The main effect of Contrast, as well as other interactions, were not significant. Pairwise comparisons for the interaction between Spatial Frequency and EFE showed that a better accuracy for happy than fearful targets was only significant when images were filtered in HSF (*M ± SD for happy faces*: .68 ± .15; *M ± SD for fearful faces*: .62 ± .12; $p_{corrected} < .001$). Furthermore, accuracy for happy targets was higher when images were in HSF ($p_{corrected} = .001$) and BSF ($M \pm SD$: .67 ± .14; $p_{corrected} = .012$) compared to LSF ($M \pm SD$: .62 ± .14) images, whereas for fearful target accuracy was marginally higher for BSF ($M \pm SD$: .66 ± .14) than HSF ($p_{corrected} = .08$) images.

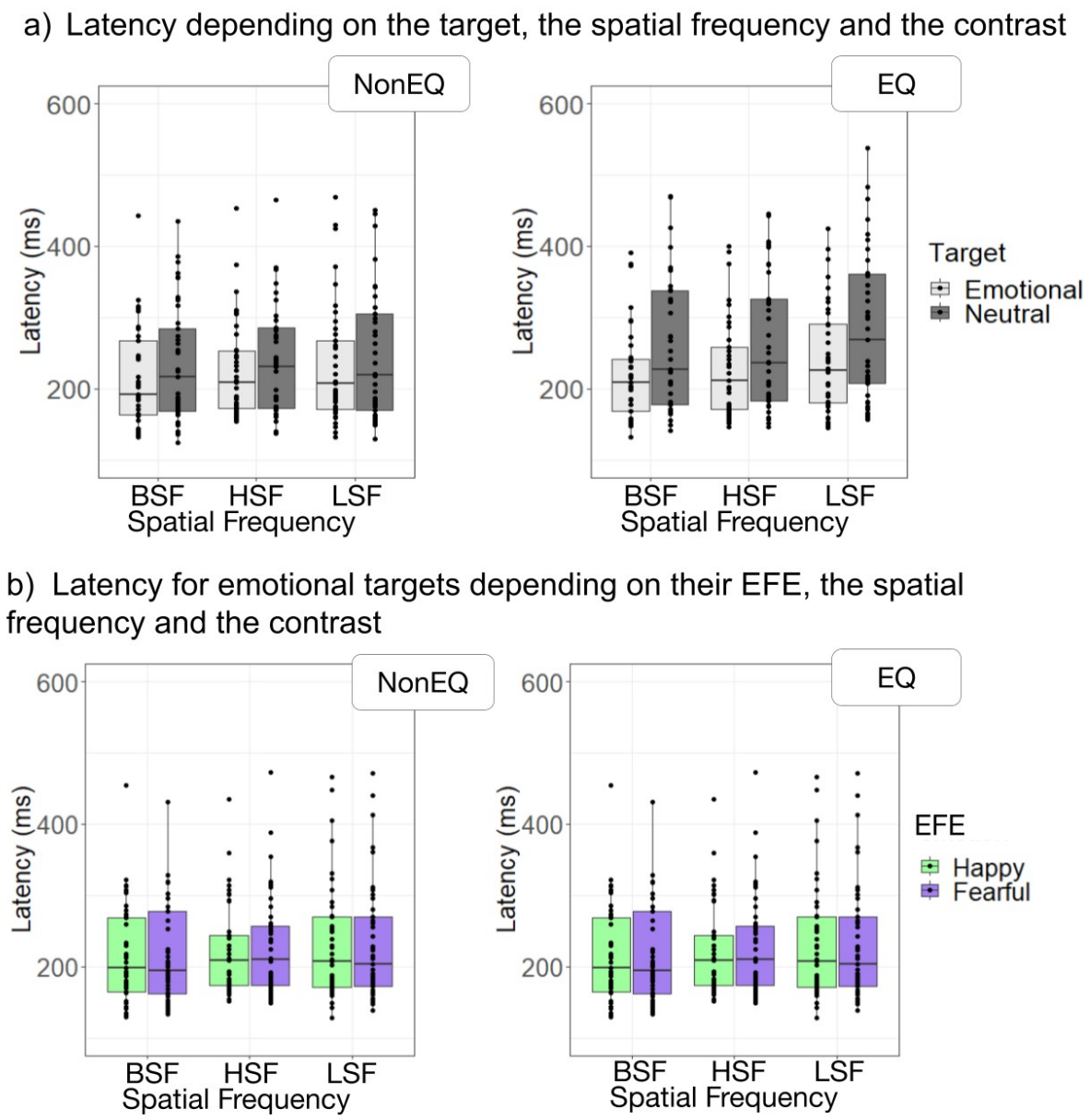

Figure 3. Boxplots for mean latency of correct saccadic responses (in ms) (a) for emotional (black) and neutral (gray) targets, for the three spatial frequency conditions (BSF, HSF, and LSF) and the two contrast conditions (NonEQ and EQ) and (b) for emotional targets, with a fearful (purple) or happy (green) face, for the three spatial frequency conditions (BSF, HSF, and LSF) and the two contrast conditions (NonEQ and EQ).

Latency: For emotional face targets, mixed ANOVA performed on mean latency of correct responses (Fig 3b) indicated only a significant main effect of the Spatial Frequency ($F(2,152)$ = 12.26, $p < .001$, $\eta_p^2 = .14$). Latency was shorter for BSF ($M \pm SD$: 217 ± 67 ms) than HSF ($M \pm SD$: 226 ± 68.4 ms; $p = .007$) or LSF ($M \pm SD$: 236 ± 79.4 ms; $p < .001$) images, as well as for HSF than LSF ($p < .001$) images. The main effect of EFE and Contrast, as well as all interactions, were not significant.

2.2.4. Analysis as a function of the EFE of the distractor for neutral face targets

Accuracy: For neutral face targets, mixed ANOVA performed on mean accuracy (Fig. 4a) indicated a marginal effect of the EFE of the distractor ($F(1,76) = 3.8$, $p = .05$, $\eta_p^2 = .047$). Performances where marginally better when the distractor was happy (*M ± SD:* .60 ± .13) compared to fearful (*M ± SD:* .58 ± .12). A significant interaction between the Spatial Frequency and the EFE of the distractor ($F(2,152) = 4.3$, $p = .016$, $\eta_p^2 = .053$) was observed. Pairwise comparisons showed that the better accuracy with happy distractors was only significant in HSF (*M ± SD for happy distractors:* .62 ± .15; *M ± SD for fearful distractors:* .57 ± .14; $p_{corrected} = .01$).

Latency: For neutral face targets, mixed ANOVA performed on mean latency (Fig 4.b) indicated a significant effect of the Spatial Frequency ($F(2,152) = 10.2$, $p < .001$, $\eta_p^2 = .11$), and, a marginal effect of the EFE of the distractor ($F(1,76) = 3.71$, $p = .057$, $\eta_p^2 = .04$). Latency was shorter for images in BSF (*M ± SD:* 246.6 ± 88.5 ms) and HSF (*M ± SD:* 249.8 ± 87.6 ms) compared to LSF (*M ± SD:* 263.6 ± 97.9 ms; $p < .001$). Latency was marginally shorter when the distractor was a happy (*M ± SD:* 252.1 ± 88.1 ms) than a fearful (*M ± SD:* 254.5 ± 90.24 ms) face. A significant interaction between the Spatial Frequency and the Contrast ($F(2,152) = 3.1$, $p = .047$, $\eta_p^2 = .04$) was observed. Pairwise comparisons showed that the shorter latencies for HSF and BSF than LSF was only significant in the EQ contrast group (*M ± SD for HSF:* 262.2 ± 95.1; *M ± SD for BSF:* 261.2 ± 94.1; *M ± SD for LSF:* 286;2 ± 105; $p_{corrected} < .001$).

2.3. Discussion

Results revealed that performances were overall better (higher mean accuracy and shorter mean latency of saccadic responses) when the target was an emotional than a neutral face. In addition, they were better for HSF than LSF images. For emotional targets, the accuracy was overall higher when the emotional face was happy than fearful, as observed in Entzmann et al. (2021) with unfiltered images. Interestingly, the difference between happy and fearful faces was only significant for HSF images, and the difference between HSF and LSF images was only significant for happy faces. Altogether, this suggests that participants mainly rely on HSF information to detect emotions, and that HSF are even more useful when the emotional face is happy. Concerning the effect of the contrast, it was not as strong as one would have expected. Thus, we did not observe the expected interactions on the accuracy (i.e., the interaction between the spatial frequencies and the contrast group, or between the EFE, the spatial frequencies and the contrast group for emotional targets). On latencies, on the other hand, the interaction between the spatial frequencies and the contrast group was marginally significant. Even this effect did not reach the significance level, it still shows that the advantage of HSF over LSF filtered faces was only significant in the EQ condition, i.e. when contrast the contrast of HSF faces is enhanced to match that of BSF and LSF images. Thus, contrast

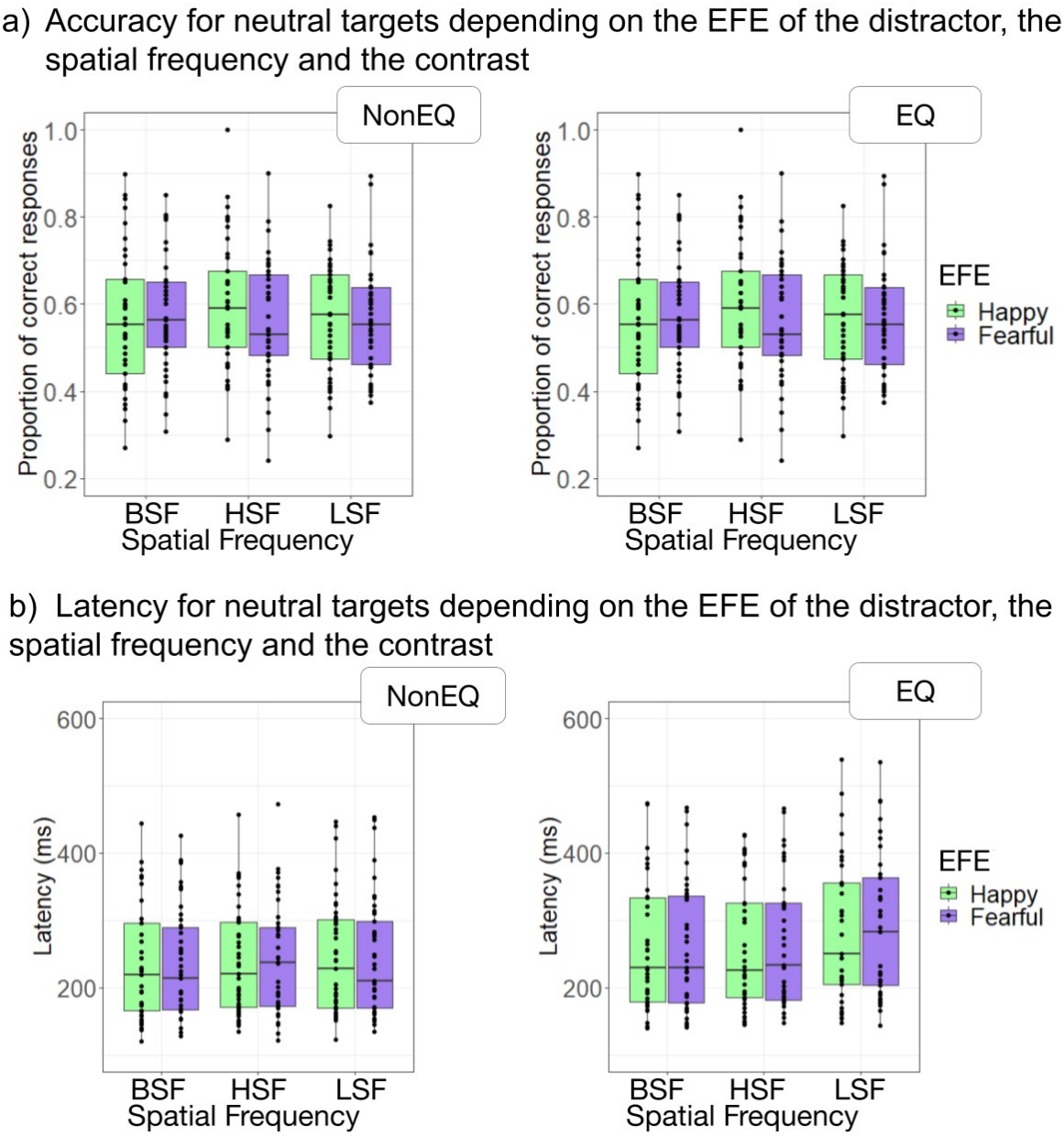

Figure 4: Boxplots for mean accuracy (a) and mean latency (b) for neutral targets, with a fearful (purple) or happy (green) distractor, for the three spatial frequency conditions (BSF, HSF and BSF) and two contrast conditions (EQ and NonEQ).

still seems to have a role in spatial frequency processing. Overall, contrary to what was expected, Furthermore, the analysis of the effect of the expression of distractor, when the target was the neutral face, allowed us to show that accuracy was higher in HSF, with happy distractors. Thus, happy HSF faces do not seem to attract the attention more, otherwise they would have caused more errors as distractors. Rather, it seems that these happy faces make the task easier, especially in HSF, by conveying the most useful information to dissociate the neutral face from the emotional face.

3. Analysis of the eye and the mouth saliency

The purpose of this analysis was to see if local statistical differences in the saliency of the mouth or the eyes, that are usually considered as diagnostic features for expression decoding, could explain participants' behaviour in the saccadic choice task. Calvo and Nummenmaa (2011) suggested in their study of happy face detection, that emotion detection may arise from a two-stage processing mechanism. The first stage would be purely perceptual (i.e., only rely on the physical attribute of the face), with an analysis of visually salient regions. In the second stage, the detection of salient features would be used for expression recognition and semantic retrieval. Visual saliency would therefore be important because it would be linked to the efficiency of the decoding. The more salient the informative regions are, the better the detection will be. Indeed, we supposed that the conditions in which the mean performance are the highest (e.g., in HSF, with a happy emotional target), are the conditions in which the diagnostic features are, on average, the most salient.

Visual saliency can be viewed as the intensity with which a region will attract attention, independently of task demands, and several computational models were proposed to compute saliency and made behavioural predictions (Borji & Itti, 2013; Foulsham & Underwood, 2008; Itti & Koch, 2000; Marat et al., 2009). The calculation of saliency differs depending on the model, but overall, a specific region is likely to be salient when it can be easily distinguished from its neighborhood, based on low-level visual attributes (Koehler et al., 2014). In their study, Calvo and Nummenmaa (2011) studied the role of perceptual (e.g., luminance, mouth or eye saliency differences), as well as higher level factors (e.g., valence) on the detection of happy faces in a saccadic choice task. They only found a contribution of the mouth saliency. The target was always a happy face, and they studied the difference between the mouth or eye saliency of the target and that of the distractor (a neutral, sad, angry, fearful, disgusted or surprised face). Here, we expected that, the more salient the mouth or eye of the target face is, the better the performances were in the saccadic choice task. Indeed, even if Calvo and Numenmaa (2011) found no contribution of the eye saliency in their happy face detection task, we still analyzed it, as it may still contribute for the detection of emotional or neutral faces. In fact, the eyes were shown to be an important feature for expressions decoding (Eisenbarth & Alpers, 2011; Wegrzyn et al., 2017), especially with fearful faces, which were used in our study (Smith & Schyns, 2009).

3.1. Method

3.1.1 Computation of the mouth and the eye saliency

The eye and mouth saliency computation were performed in two steps. First, saliency maps were generated (i.e., on happy, neutral, or fearful faces, each in BSF, HSF, or LSF, with or without contrast equalization), then rectangular boxes encompassing the eyes and mouth were used to compute the average saliency of the eyes and mouth for each saliency map. Maps were computed for each image using the computational saliency model proposed by Walther and Koch (2006). This model is inspired by the biology of the human visual system and considers intensities and orientations to attribute a saliency value to each pixel of an input image. The model was implemented in Matlab using the SaliencyToolbox (http://www.saliencytoolbox.net) and the *makeSaliencyMap* function. Note that we have changed the default settings for calculating the maps. First the center-surround parameters were lowered to allow the identification of more localized regions. Then, the maps were generated after a single iteration, as we were not interested in any inhibition of return mechanism (See Appendix A for an overview of the used parameters). At the end of this process, a saliency map was obtained for each image. Figure 5a shows examples of saliency maps superimposed to their corresponding stimuli (i.e., a specific face in the different spatial frequency, contrast and emotion conditions).

The mouth and eye regions were then visually selected, so that we obtained rectangular boxes that contains the mouth or the eyes (including the eyebrows) for all our individuals, with any expression. More precisely, the mouth region corresponded to pixels with a X-coordinate between 95 and 210, and a Y-coordinate between 207 and 280 (for images that sized 300x300 pixels; X_0 and Y_0 coordinates being the top-left corner). The eye region corresponded to pixels with a X-coordinate between 70 and 237, and a Y-coordinate between 110 and 172. Then the mouth and eye regions were the same for all the stimuli. Afterwards, we computed for each image the mean saliency value, by averaging all the saliency values of all the pixels contained in the box surrounding the mouth or the eyes. The mouth and eye region can be visualized in Fig.5b.

3.1.2 Data analysis

First, A factorial 3x3x2 ANOVA with the EFE (Fearful, Neutral, Happy), the Spatial Frequency (BSF, HSF, LSF) and the Contrast (NonEQ, EQ) as between-image factors was applied on the mean saliency of each image, once for the eyes and once for the mouth. Note that BSF images are the same in both contrast conditions, as contrast equalization only modulate the contrast of HSF and LSF images. We still used t-tests with Bonferroni corrections for pairwise comparisons.

Then, we also performed Pearson's correlations to evaluate the strength of the relationship between mean accuracy and latency in each condition in the saccadic choice task (independently of participants), and the mean mouth or eye saliency in the same condition (when the target was a happy, neutral or fearful face, each in BSF, HSF, or LSF, in the NonEQ or EQ contrast group). The

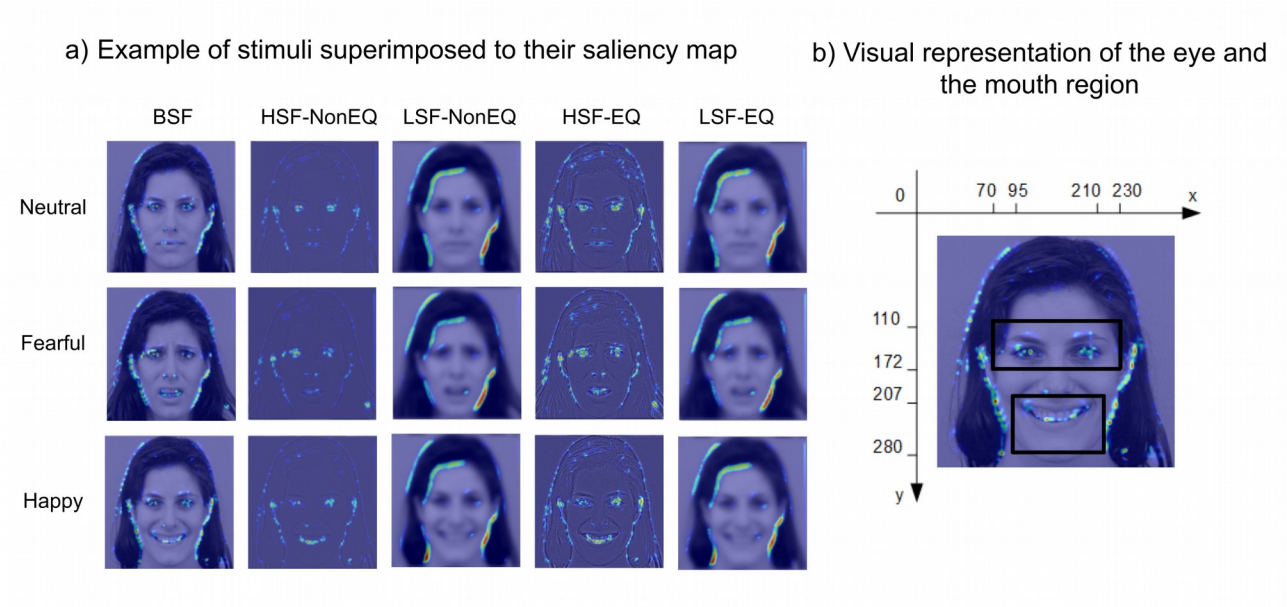

Figure 5. (a) Example of saliency maps computed for one face viewed in its different spatial frequency and contrast conditions, with a neutral, happy or fearful expression with the model from Walther and Koch (2006). Experimental conditions are (from left to right) BSF, HSF NonEQ, LSF NonEQ, HSF EQ and LSF EQ. (b) Visual representation of the eye and the mouth regions added on a saliency map. These boxes were used to calculate the eye and mouth saliency, by taking the average saliency value of all the pixels included in a box. For illustration purposes, we have displayed the saliency maps overlaid on the stimuli from which they were computed.

goal was to determine if the conditions for which the mouth or the eyes is, on average, the most salient, were the conditions in which the performance were, on average, the highest. Therefore, for each correlation, we studied the relationship between 18 saliency values and 18 performance values (1 value for each condition: when the target was a happy, neutral or fearful face, each in BSF, HSF, or LSF, in the NonEQ or EQ contrast group).

3.2. Results

3.2.1 Eye saliency

The ANOVA performed on mean eye saliency (Fig.6a) showed a significant main effect of the EFE ($F(2, 342) = 6.08$, $p = .002$, $\eta_p^2 = .034$), the Spatial Frequency ($F(2, 342) = 5.44$, $p = .005$, $\eta_p^2 = .031$), and the Contrast ($F(1, 342) = 11.6$, $p < .001$, $\eta_p^2 = .033$). Overall, mean eye saliency was higher for neutral ($M \pm SD: 0.0355 \pm 0.014$) than happy ($M \pm SD: 0.03 \pm 0.014$; $p = .003$) and fearful ($M \pm SD: 0.031 \pm 0.012$; $p = .007$) faces. It was also higher for BSF ($M \pm SD: 0.035 \pm 0.013$; $p = .003$) and HSF ($M \pm SD: 0.033 \pm 0.013$; $p = .032$) than LSF ($M \pm SD: 0.029 \pm 0.013$) images. Finally, it was overall higher for the EQ ($M \pm SD: 0.035 \pm 0.014$) than NonEQ ($M \pm SD:$

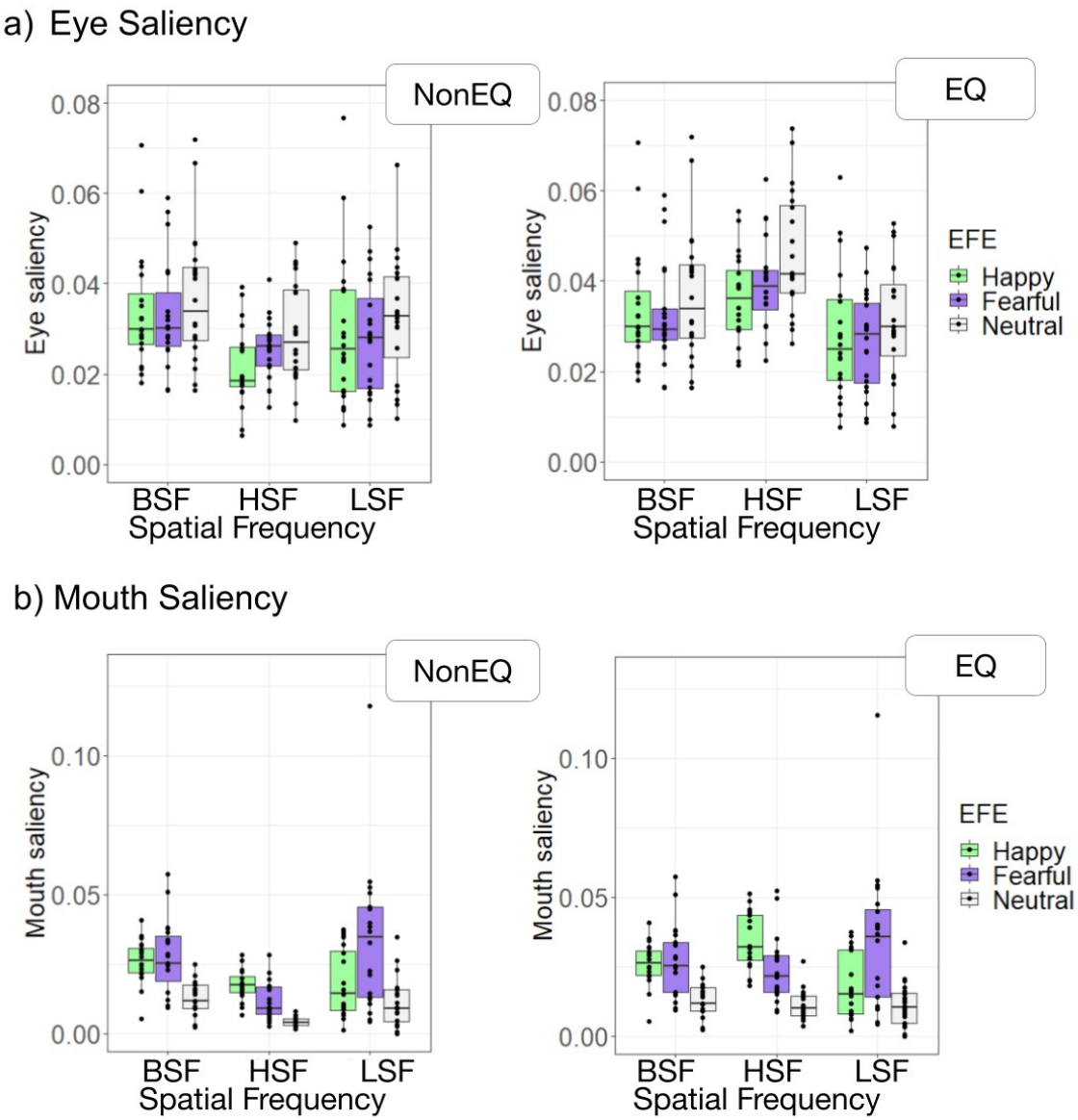

Figure 6. Boxplots for mean (a) eye or (b) mouth saliency, for happy (green), fearful (purple) or neutral (gray) faces, for the three spatial frequency conditions (BSF, HSF, and LSF) and the two contrast conditions (NonEQ and EQ).

0.03 ± 0.013) contrast condition. Moreover, there was a significant interaction between the Spatial Frequency and the Contrast ($F(2,342) = 15.9$, $p < .001$, $\eta_p^2 = .085$). All other interactions were not significant.

Pairwise comparison for the interaction between the Spatial Frequency and the Contrast revealed that eye saliency was higher for EQ than NonEQ stimuli filtered in HSF only (*M ± SD for the NonEQ condition: 0.025 ± 0.01; M ± SD for the EQ condition: 0.041 ± 0.012*). Furthermore, in the NonEQ condition, eye saliency was higher for BSF (*M ± SD: 0.034 ± 0.013*) than HSF ($p_{corrected} < .001$) images, whereas in the EQ condition it was higher for HSF than LSF (*M ± SD: 0.029 ± 0.013;* $p_{corrected} < .001$) images.

3.2.2 Mouth saliency

The ANOVA performed on mean mouth saliency (Fig.6b) revealed a significant main effect of the EFE ($F(2, 342) = 55.9$, $p < .001$, $\eta_p^2 = .25$), the Spatial Frequency ($F(2, 342) = 5.77$, $p = .003$, $\eta_p^2 = .032$), and the Contrast ($F(1, 342) = 8.81$, $p = .003$, $\eta_p^2 = .025$). Overall, mouth saliency was higher for happy (*M ± SD: 0.023 ± 0.011)* and fearful (*M ± SD: 0.026 ± 0.019*) than neutral faces (*M ± SD: 0.011 ± 0.001*; *p < .001*). It also tended to be higher for fearful than happy faces ($p = 0.09$). Then, it was higher for BSF (*M ± SD: 0.022 ± 0.011; p =.008*) and LSF (*M ± SD: 0.021 ± 0.011; p =.038*) than HSF (*M ± SD: 0.017 ± 0.012)* images. Concerning the contrast equalization, mean mouth saliency was higher in the EQ (*M ± SD: 0.022 ± 0.015)* than in the NonEQ (*M ± SD: 0.018 ± 0.15)* contrast condition. Moreover, there was a significant interaction between the Spatial Frequency and the EFE ($F(4,342) = 9.86$, $p < .001$, $\eta_p^2 = .1$) and between the Spatial Frequency and the Contrast ($F(2,342) = 9.8$, $p < .001$, $\eta_p^2 = .054$). All other interactions were not significant.

For the interaction between the Spatial Frequency and the EFE, pairwise comparison showed that, for LSF images mouth saliency was higher for fearful (*M ± SD: 0.034 ± 0.026)* than happy (*M ± SD: 0.018 ± 0.012)* or neutral (*M ± SD: 0.011 ± 0.008;* $p_{corrected} < .00$1) faces. For HSF images, mouth saliency was higher for happy (*M ± SD: 0.25 ± 0.012;* $p_{corrected} < .001$*)* and fearful (*M ± SD: 0.017 ± 0.011;* $p_{corrected} = .017$*)* than neutral (*M ± SD: 0.007 ± 0.005)* faces. Similarly, for BSF images, it was higher for happy (*M ± SD: 0.26 ± 0.008)* and fearful (*M ± SD: 0.027 ± 0.013)* than neutral (*M ± SD: 0.013 ± 0.006;* $p_{corrected} < .001$*)* faces. For happy and neutral faces, there was no differences between frequency conditions. For fearful faces, mouth saliency was higher for LSF ($p_{corrected} <.001$) and BSF ($p_{corrected} = .024$) than HSF images.

Concerning the interaction between the Contrast and the Spatial Frequency, the mouth saliency was higher for the EQ than NonEQ condition for HSF filtered faces only (*M ± SD for the NonEQ condition: 0.011 ± 0.007; M ± SD for the EQ condition: 0.023 ± 0.013;* $p_{corrected} = .003$). In the NonEQ condition, mouth saliency was higher for BSF (*M ± SD: 0.022 ± 0.011;* $p_{corrected} < .001$ *)* and LSF (*M ± SD: 0.021 ± 0.02;* $p_{corrected} = .003$*)* than HSF images, whereas there was no difference between spatial frequency conditions in the EQ group.

3.2.3 Relationship between saliency and behavioural data

***Accuracy*:** Mean accuracy in the saccadic choice task in each condition was found to be positively correlated with the mean mouth saliency of the target ($r = .72$, $p < .001$). The more the mouth of the target was salient, the better the accuracy. There was no significant correlation between mean accuracy in the saccadic choice task and the eye saliency ($r = .14$, $p = .57$). The proportion of correct saccades as a function of the mouth or eye saliency of the target is presented in Fig.7a, in which one point corresponds to one condition.

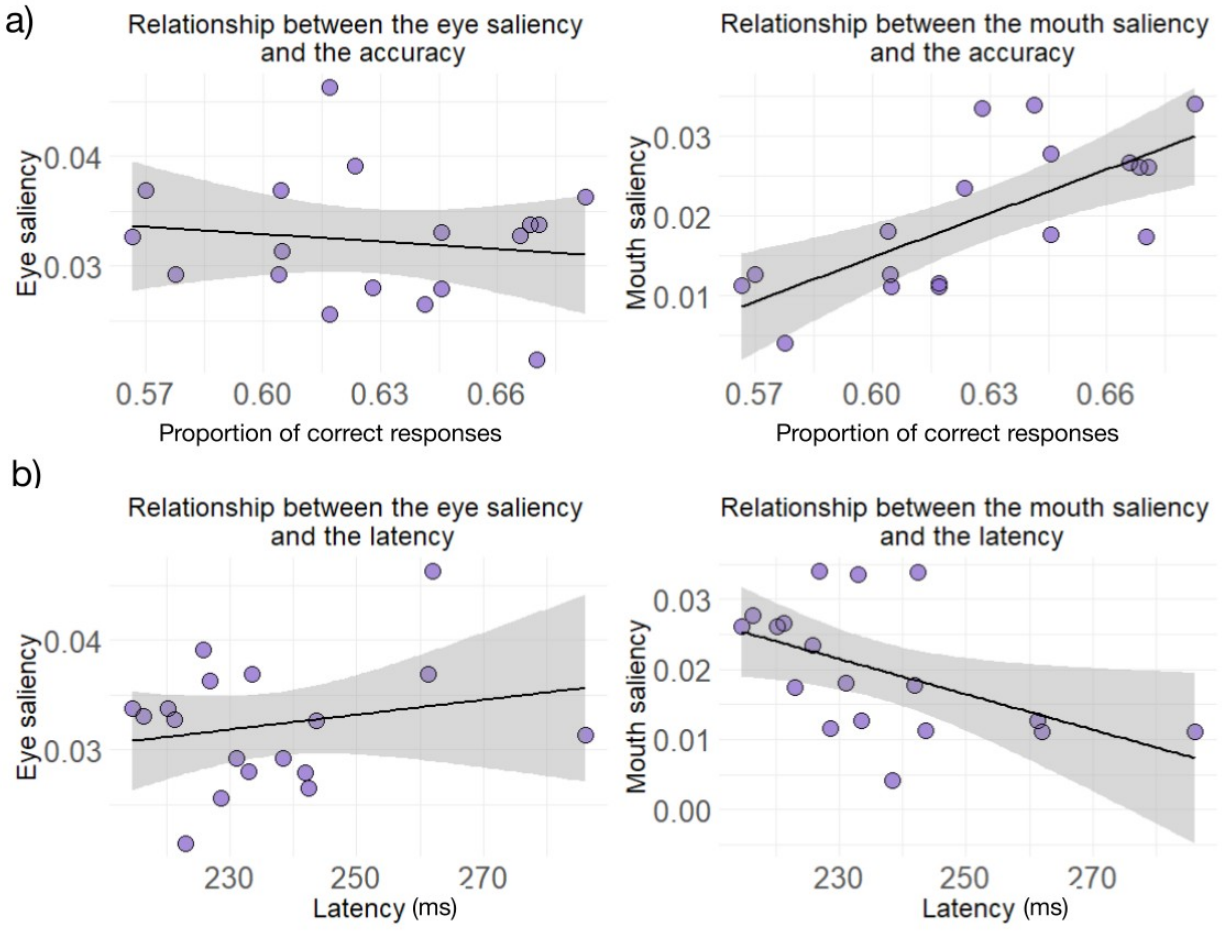

Figure 7. Relationship between (a) the accuracy or (b) the latency and the eye (left) or mouth (right) saliency. One dot corresponds to one condition of the saliency analysis (a happy, fearful, or neutral target, each in BSF, HSF, or LSF, with or without contrast equalization).

***Latency*:** Mean latency in the saccadic choice task in each condition was found to be negatively correlated with the mean mouth saliency of the target (r = -.5, p = .033). The higher the mouth was salient, the shorter the latency. There was no significant correlation between mean latency in the saccadic choice task and the eye saliency (r = .22, p = .39). The saccade latency as a function of the mouth or eye saliency of the target is presented in Fig.7b.

3.3. Discussion

Overall, the mouth saliency was found to correlate significantly with behavioural results, whereas the eye saliency does not seem to be very important for this task. Correlation analysis shows that the greater the mouth saliency of the target was in one condition, the better the performance was (higher accuracy, shorter latency). Especially, while looking at the mean mouth saliency in the different conditions, we can see that it is higher for emotional than neutral faces. Furthermore, mean mouth saliency was also higher for happy than fearful faces in HSF, and for

HSF than LSF for happy faces. Although this effect did not reach the significant level in the statistical analysis after the Bonferroni correction was applied, the direction of the differences is still in line with the behavioural results on the accuracy. Finally, the fact that the mouth saliency was higher for HSF in the EQ than NonEQ condition is in line with the results on the latency of participants, showing that the HSF over LSF advantage was only significant in the EQ condition. Overall, we showed that there is a link between the mouth saliency and the performance. However, the differences are not systematically the same. For example, the mouth of LSF fearful faces in the NonEQ contrast condition is particularly salient, and this is not well represented in the participant accuracy. Thus, even if the mouth saliency is related to the participant's performance, it is still likely that other mechanisms are involved.

4. General discussion

The aim of this study was to clarify the role of spatial frequency and luminance contrast in the detection of emotional faces through an eye-tracking experiment and an analysis of the saliency of the diagnostic face features (the eyes and the mouth) of the stimuli. For the eye-tracking experiment, based on a saccadic choice task, we replicated findings from previous studies. Performances were greater (higher accuracy, shorter latency) when the target was an emotional than a neutral face (Bannerman et al., 2009; Entzmann et al., 2021). For emotional targets, accuracy was higher when the emotional face was happy compared to fearful; and this effect was mainly driven by HSF. Also, we found that the accuracy was overall higher for HSF and BSF than LSF images; The HSF over LSF advantage was mainly driven by happy face targets. Concerning the latencies, they were overall shorter for BSF than filtered images, and for HSF compared to LSF images. Finally, the impact of contrast equalization on performances was not as high as we expected. Nevertheless, the marginal interaction between the spatial frequencies and the contrast that we observed on saccade latencies showed that the difference between HSF and LSF was significant for equalized images only. This suggest that the processing of spatial frequencies can be dependent on contrast differences, which would, at least partially, explain differences in the results of neuroimaging studies (McFadyen et al., 2017).

To explain our behavioural results, we analyzed in the second part of the paper, the saliency maps associated with our stimuli. As several previous studies suggest that the eyes and the mouth are the most important regions for expression decoding (Eisenbarth & Alpers, 2011; Smith & Schyns, 2009; Sweeny et al., 2013; Wegrzyn et al., 2017), we focused in our analyzes on the saliency of the mouth and the eye regions. This saliency analysis was also motivated by some of our previous findings (Entzmann et al., 2021). Indeed, using similar saccadic choice tasks (i.e., with

emotional faces; either presented in emotional-neutral or face-vehicle pairs) we showed that saccades landed differently on the faces according to the expression (e.g., saccades were lower on emotional than neutral faces, especially happy ones). We supposed that the saliency of the face features differentiating each expression can modulate the attention, leading to differences in saccade endpoints. Here, we were specifically interested in how this can be linked with the performance of participants. Overall, in the present study, unlike the mean eye saliency, the mean mouth saliency in the different experimental conditions was found to correlate with both the mean accuracy, and the mean latency of participants.

A primary use of HSF despite the statistical sufficiency of LSF

According to the coarse-to-fine theory of visual processing, LSF information is processed faster than HSF information (Bar, 2003; Hegde, 2008; Kauffmann et al., 2014; Kauffmann, Chauvin, et al., 2015; Musel et al., 2012; Peyrin et al., 2010; Schyns & Oliva, 1994). However, although results of our saccadic choice task showed that LSF information is sufficient to accurately detect an emotion, as accuracy with LSF images was above chance, we observed shorter latencies for HSF. A possible explanation for the HSF bias in the saccadic choice task is that participants favor HSF because LSF are not informative enough to rapidly disambiguate the emotional or neutral content of faces. In our task we have two faces side by side, which probably have the same global structure (thus a similar LSF content). To differentiate between the two in terms of emotion, we would have to rely on more detailed information. Indeed, this is in line with the idea that there is a flexible use of spatial frequencies for facial expression decoding. Different frequencies would be extracted depending on the expression (Morrison & Schyns, 2001; Oliva & Schyns, 1997; Schyns et al., 2009; Smith et al., 2005) and the task (Schyns & Oliva, 1999; Smith & Merlusca, 2014). For example, Schyns and Oliva found that when participants were asked to categorize facial expressions as angry, happy or neutral, they relied more on LSF information whereas they relied on HSF when they had to indicate whether the face was expressive or neutral (Schyns & Oliva, 1999), a result that is consistent with our data. Therefore, the use of spatial frequency information would not be a fixed process, as it may depend on the scale of the local features that are diagnostic, considering both task constraints and the spatial configuration of the face.

In our task, we showed that the mouth saliency is correlates with performances, but there may be other processes that contribute to the HSF over LSF preference. In this task we can think that participants attend for local more than global difference between the two face images which are structurally similar. Several studies show that the attention can flexibly be directed to both global and local levels depending on expectations. For example, studies using Navon display showed enhance processing of HSF after attending to local structure, and enhance processing of LSF after

attending to global structure (Ivry and Robertson 1998; Robertson and Ivry 2000; Flevaris et al. 2011). Moreover, some previous studies also proposed that stimulus duration influences the use of spatial information, and, that there is a HSF over LSF bias for long stimulus duration (Peyrin et al., 2006; Mermillod et al., 2010; Schyns & Oliva 1994). A well-known study from Schyns and Oliva (1994) used hybrid stimuli with two natural scene images from different categories and frequency scale to test the influence of stimuli duration. They found that very short presentation time (30 ms) elicited a categorization based on the LSF content whereas longer presentation time (150 ms) elicited a categorization based on the HSF content. In our task stimulus where displayed for a long presentation time (800 ms), there was therefore no strong time constraint, which could explain the HSF bias. Nevertheless, in saccadic choice tasks with face-vehicle pairs, a LSF bias was found even with a relatively long presentation time (400 ms; Guyader et al., 2017). This could suggest that if the LSF content is distinct enough between the target and distractor (like with faces and vehicle that would be easy to distinguish in LSF) it could be favored regardless of the presentation time.

Overall, contrary to our hypothesis, we found no evidence for a better detection of LSF fearful faces, suggesting that they are not automatically (i.e., quickly and unintentionally) prioritized compared to other emotional faces competing for attention. Such a hypothesis (i.e., better detection of LSF fearful faces) was mainly based on neurophysiological data showing enhanced, or earlier, amygdala activation for LSF fearful faces, and on the existence of a subcortical pathway for rapid threat detection (LeDoux, 2000; Morris et al., 1999; Öhman, 2005; Tamietto & de Gelder, 2010; Vuilleumier et al., 2003; Méndez-Bértolo et al., 2016). However, it is important to clarify that the objective of the present study was not to provide evidence for such a pathway, a behavioural study alone would not be conclusive without being coupled with neuroimaging techniques. Rather, the objective was to test if, at the behavioural level, these stimuli could be detected more efficiently even when opposed to other faces. In this sense, we cannot exclude that LSF fearful faces activated the amygdala earlier in this task, but we showed that such an effect was not reflected here on eye movements. Then, even if a strict automaticity of the prioritization of LSF fearful faces is unlikely considering our results, it is still possible that it exists in other tasks. For example, while neurophysiological data diverge on the effect the orientation of the attention on amygdala activity (Bayle & Taylor 2010; Pessoa et al., 2002; Habel et al., 2007; Vuilleumier and Schwartz, 2001; Whalen et al.,1998), a prioritization of LSF fearful faces could still arise for implicit detection, as here we only explicitly assessed emotion detection. For example, this could be assessed if we replicate the experiment but without instructions on the target to see where the saccades go first.

The role of contrast equalization and local face features on the detection of emotions

The particularity of this study was also to consider contrast differences between HSF and LSF, by comparing performance in the saccadic choice task and in the mouth or eye saliency across two contrast conditions: one in which the RMS contrast was equalized after the filtering process, and one in which it was not (i.e., LSF images had a higher luminance contrast than HSF images). In the behavioural experiment, the expected interaction between contrast and spatial frequency was marginally significant on saccade latencies. More precisely, the difference between HSF and LSF conditions was only significant with equalized images. Even if the interaction did not reach the significance level, the direction of the effect is in line with previous findings (Vlamings et al., 2009; Perfetto et al., 2020; Kauffmann, Chauvin, et al., 2015; Kauffmann, Ramanoël, et al., 2015). For example, Kauffmann, Chauvin et al. (2015) observed that scene categorization as indoor or outdoor in HSF was slower without contrast equalization than with contrast equalization. In another study performed by Vlamings et al. (2009) using LSF or HSF faces, with or without contrast equalization, participants were asked to decide whether the presented stimulus was a fearful or a neutral face. Results showed that LSF faces were categorized more rapidly than HSF faces, for both equalized and non-equalized stimuli. However, this observed difference was stronger when contrast was not equalized between LSF and HSF faces. Overall, this better processing of HSF after contrast equalization is also congruent with the saliency analysis showing enhanced mouth or eye saliency in HSF when the contrast is equalized. The fact that the spatial frequency and contrast interaction was only marginal and not significant interaction on saccade latencies could be due to a lack of power. Also, the effect size may be smaller in this study because the spatial frequency conditions were presented in blocks, allowing participant to get used to the current contrast and take it as a baseline. Saccade accuracy was not sensitive to different contrast conditions, suggesting that overall contrast is not important for the accuracy of a discrimination, which may be mostly based on within-pairs differences.

Both saccade accuracy and latency correlate with the mouth saliency, supposing that local contrasts and orientations in the diagnostic regions contribute to the efficiency of the detection of emotional and neutral faces. Interestingly, whereas the mouth saliency was important for participants, the eye saliency didn't influence the performance. This idea that the eye region plays a limited role in the detection of neutral and emotional faces contrasts with studies showing that both the eye and the mouth are important (Dailey et al., 2002; Eisenbarth & Alpers, 2011; Smith & Schyns, 2009; Wegrzyn et al., 2017). However, such studies focused on the categorization of emotions. They showed that the mouth is particularly useful for the categorization of happy faces and the eyes for fearful faces. Here, the task is different, and closer to a categorization of faces as emotional or neutral. Therefore, it is not so surprising that the features are used differently than when the task is to categorize precisely the expression as happy or fearful. It is also is possible that

participants make the strategy to focus on the mouth because they consider that it is the most informative region. Going further we can suppose that such a focus would favor the detection of emotional faces when they are happy. Indeed, for happy faces the mouth region is highly diagnostic, whereas for fearful faces it is less stereotypical. Finally, our results are limited to the use of happy and fearful faces. We can suppose that with other emotional faces the results could have been different, especially if the emotional faces convey useful information through the eye region (e.g., fearful and angry faces; Smith et al., 2005).

5. Conclusion

This study replicated findings from previous studies showing a better detection of emotional than neutral faces, especially with a happy emotional face (Entzmann et al., 2021). Also, we found that the discrimination between an emotional and a neutral face was overall easier for HSF and BSF than LSF images. On saccade latency, the HSF over LSF advantage was significant for equalized images only, suggesting that the processing of spatial frequencies can be dependent on contrast differences. The saliency analysis of our stimuli revealed that the mean mouth saliency in the different experimental conditions, unlike the eye saliency, correlated with both the mean accuracy and the mean latency of participants. Overall, our results go against the idea that there is an automatic (quick and unintentional) prioritization of low spatial frequency fearful faces. Rather, we suggest that participants favoured the use of high spatial frequencies in this task because low spatial frequencies are not informative enough to rapidly disambiguate the emotional or neutral content of faces. This would be consistent with the idea that there is a flexible use of spatial frequencies for facial expression decoding, depending of the scale of the useful information for a specific task (Smith & Merlusca, 2014). Also, as suggested in previous papers, the saliency of diagnostic features, the mouth in this study, may be used as a shortcut for efficient expression decoding (Calvo & Nummenmaa, 2016).

6. Acknowledgments

This work was supported by NeuroCoG IDEX UGA in the framework of the “Investissements d’avenir” program (ANR-15-IDEX-02). This work has been partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003).

7. Conflict of interest

The authors declare no conflict of interest.

8. Data availability statement

The data that support the findings of this study are available in the Open Science Framework repository at https://osf.io/hyr52.

9. References

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. https://doi.org/10.1038/nature03086

Awasthi, B., Friedman, J., & Williams, M. A. (2011). Faster, stronger, lateralized : Low spatial frequency information supports face processing. *Neuropsychologia*, 49(13), 3583-3590. https://doi.org/10.1016/j.neuropsychologia.2011.08.027

Bannerman, R. L., Milders, M., & Sahraie, A. (2009). Processing emotional stimuli: Comparison of saccadic and manual choice-reaction times. *Cognition & Emotion*, *23*(5), 930–954. https://doi.org/10.1080/02699930802243303

Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, 15(4), 600-609. https://doi.org/10.1162/089892903321662976

Bayle, D. J., & Taylor, M. J. 2010. Attention inhibition of early cortical activation to fearful faces. *Brain Research*, 1313: 113–123.

Borji, A., & Itti, L. (2013). State-of-the-Art in Visual Attention Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 185-207. https://doi.org/10.1109/TPAMI.2012.89

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433-436. https://doi.org/10.1163/156856897X00357

Calvo, M. G., & Nummenmaa, L. (2009). Eye-movement assessment of the time course in facial expression recognition: Neurophysiological implications. *Cognitive, Affective, & Behavioral Neuroscience*, 9(4), 398–411. https://doi.org/10.3758/CABN.9.4.398

Calvo, M. G., & Nummenmaa, L. (2011). Time course of discrimination between emotional facial expressions : The role of visual saliency. *Vision Research*, 51(15), 1751-1759. https://doi.org/10.1016/j.visres.2011.06.001

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. Cognition and Emotion, 30(6), 1081-1106.

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task : Experimental data and computer model. Journal of Vision, 9(12), 10-10. https://doi.org/10.1167/9.12.10

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, 14(8), 5. https://doi.org/10.1167/14.8.5

Crouzet, S. M. (2010). Fast saccades toward faces : Face detection in just 100 ms. *Journal of Vision*, 10(4), 1-17. https://doi.org/10.1167/10.4.16

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A Neural Network that Categorizes Facial Expressions. *Journal of Cognitive Neuroscience*, 14(8): 1158–1173.

Devue, C., & Grimshaw, G. M. (2017). Faces are special, but facial expressions aren't: Insights from an oculomotor capture paradigm. *Attention, Perception, & Psychophysics*, *79*(5), 1438–1452. https://doi.org/10.3758/s13414-017-1313-x

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes : Scanning emotional facial expressions. *Emotion*, 11(4), 860-865. https://doi.org/10.1037/a0022758

Entzmann, L., Guyader, N., Kauffmann, L., Lenouvel, J., Charles, C., Peyrin, C., Vuillaume, R. and Mermillod, M. (2021), The Role of Emotional Content and Perceptual Saliency During the Programming of Saccades Toward Faces. Cognitive Science, 45: e13042. https://doi.org/10.1111/cogs.13042

Farah, M. J., Wilson, K. D., & Drain, M. (1998). What Is « Special » About Face Perception? 17.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America* A, 4(12), 2379. https://doi.org/10.1364/JOSAA.4.002379

Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat : Amygdala reactivity across multiple expressions of facial affect. *NeuroImage*, 30(4), 1441-1448. https://doi.org/10.1016/j.neuroimage.2005.11.003

Flevaris, Anastasia V., Bentin, S., & Robertson, L. C. (2011). Attention to hierarchical level influences attentional selection of spatial scale. *Journal of Experimental Psychology: Human Perception and Performance*, 37(1): 12–22.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, 8(2), 6. https://doi.org/10.1167/8.2.6

Garvert, M. M., Friston, K. J., Dolan, R. J., & Garrido, M. I. (2014). Subcortical amygdala pathways enable rapid face processing. *NeuroImage*, 102, 309-316. https://doi.org/10.1016/j.neuroimage.2014.07.047

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From Coarse to Fine? Spatial and Temporal Dynamics of Cortical Face Processing. *Cerebral Cortex*, 21(2), 467-476. https://doi.org/10.1093/cercor/bhq112

Goffaux, V., & Rossion, B. (2006). Faces are « spatial »—Holistic face perception is supported by low spatial frequencies. Journal of Experimental Psychology: *Human Perception and Performance*, 32(4), 1023-1039. https://doi.org/10.1037/0096-1523.32.4.1023

Guyader, N., Chauvin, A., Boucart, M., & Peyrin, C. (2017). Do low spatial frequencies explain the extremely fast saccades towards human faces? *Vision Research*, 133, 100-111. https://doi.org/10.1016/j.visres.2016.12.019

Habel, U., Windischberger, C., Derntl, B., Robinson, S., Kryspin-Exner, I., Gur, R. C., & Moser, E. (2007). Amygdala activation and facial expressions: explicit emotion discrimination versus implicit emotion processing. *Neuropsychologia*, *45*(10), 2369-2377.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223-233. https://doi.org/10.1016/S1364-6613(00)01482-0

Hegde, J. (2008). Time course of visual perception : Coarse-to-fine processing and beyond. Progress in *Neurobiology*, 84(4), 405-439. https://doi.org/10.1016/j.pneurobio.2007.09.001

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12), 1489-1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Ivry, R. B., Robertson, L. C., & Robertson, L. C. (1998). *The two sides of perception*. MIT press.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, 6(10), 766-774. https://doi.org/10.1038/nrn1766

Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization : Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, 107, 49-57. https://doi.org/10.1016/j.visres.2014.11.013

Kauffmann, L., Khazaz, S., Peyrin, C., & Guyader, N. (2021). Isolated face features are sufficient to elicit ultra-rapid and involuntary orienting responses toward faces. *Journal of Vision*, 21(2), 4. https://doi.org/10.1167/jov.21.2.4

Kauffmann, L., Peyrin, C., Chauvin, A., Entzmann, L., Breuil, C., & Guyader, N. (2019). Face perception influences the programming of eye movements. *Scientific Reports*, 9(1). https://doi.org/10.1038/s41598-018-36510-0

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, 8. https://doi.org/10.3389/fnint.2014.00037

Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, 112, 86-95. https://doi.org/10.1016/j.neuroimage.2015.02.058

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements : Visual processing speed revisited. *Vision Research*, 46(11), 1762-1776. https://doi.org/10.1016/j.visres.2005.10.002

Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of Vision*, 14(3), 14-14. https://doi.org/10.1167/14.3.14

Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, 107(1), 330-342. https://doi.org/10.1016/j.cognition.2007.07.012

LeDoux, J. E. (2000). *Emotion Circuits in the Brain*. 31.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception : An MEG study. *Nature Neuroscience*, 5(9), 910-916. https://doi.org/10.1038/nn909

Ludwig, C. J., Gilchrist, I. D., & McSorley, E. (2004). The influence of spatial frequency and contrast on saccade latencies. *Vision research*, *44*(22), 2597-2604.

Lundqvist, D., Flykt, A., & Ohman, A. (1998). The Karolinska directed emotional faces (KDEF).

Marat, S., Ho Phuoc, T., Granjon, L., Guyader, N., Pellerin, D., & Guérin-Dugué, A. (2009). Modelling Spatio-Temporal Saliency to Predict Gaze Direction for Short Videos. *International Journal of Computer Vision*, 82(3), 231-243. https://doi.org/10.1007/s11263-009-0215-3

McFadyen, J., Mermillod, M., Mattingley, J. B., Halász, V., & Garrido, M. I. (2017). A Rapid Subcortical Amygdala Route for Faces Irrespective of Spatial Frequency and Emotion. *The Journal of Neuroscience*, 37(14), 3864-3874. https://doi.org/10.1523/JNEUROSCI.3525-16.2017

Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y. H., Vuilleumier, P., Gil-Nagel, A., & Strange, B. A. (2016). A fast pathway for fear in human amygdala. *Nature Neuroscience*, 19(8), 1041-1049. https://doi.org/10.1038/nn.4324

Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions : Evidence from neural computation. *Neurocomputing*, 73(13-15), 2522-2531. https://doi.org/10.1016/j.neucom.2010.06.002

Mermillod, M., Vuilleumier, P., Peyrin, C., Alleysson, D., & Marendaz, C. (2009). The importance of low spatial frequency information for recognising fearful facial expressions. *Connection Science*, 21(1), 75-83. https://doi.org/10.1080/09540090802213974

Morris, J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain*, 121(1), 47-57. https://doi.org/10.1093/brain/121.1.47

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, 8(3), 454-469. https://doi.org/10.3758/BF03196180

Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is Coarse-to-Fine Strategy Sensitive to Normal Aging? *PLoS ONE*, 7(6), e38493. https://doi.org/10.1371/journal.pone.0038493

Öhman, A. (2005). The role of the amygdala in human fear : Automatic detection of threat. *Psychoneuroendocrinology*, 30(10), 953-958. https://doi.org/10.1016/j.psyneuen.2005.03.019

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. https://doi.org/10.1006/cogp.1997.0667

Perfetto, S., Wilder, J., & Walther, D. B. (2020). Effects of Spatial Frequency Filtering Choices on the Perception of Filtered Images. *Vision*, 4(2), 29. https://doi.org/10.3390/vision4020029

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala : From a « low road » to « many roads » of evaluating biological significance. *Nature Reviews Neuroscience*, 11(11), 773-782. https://doi.org/10.1038/nrn2920

Peters, J. C., Goebel, R., & Goffaux, V. (2018). From coarse to fine: Interactive feature processing precedes local feature analysis in human face perception. *Biological psychology*, *138*, 1-10.

Petras, K., Ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, *186*, 103-112.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., Marendaz, C., & Vuilleumier, P. (2010). The Neural Substrates and Timing of Top–Down Processes during Coarse-to-Fine Categorization of Visual Scenes : A Combined fMRI and ERP Study. *Journal of Cognitive Neuroscience*, 22(12), 2768-2780. https://doi.org/10.1162/jocn.2010.21424

Peyrin, C., Mermillod, M., Chokron, S., & Marendaz, C. (2006). Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing. *Brain and Cognition*, *62*(3), 214-220.

Quek, G. L., Liu-Shuang, J., Goffaux, V., & Rossion, B. (2018). Ultra-coarse, single-glance human face detection in a dynamic visual stream. *NeuroImage*, *176*, 465-476.

Racine, J. S. (2012). RStudio : A platform-independent IDE for R and Sweave. Journal of Applied Econometrics,.

Robertson, L. C., & Ivry, R. (2000). Hemispheric Asymmetries: Attention to Visual and Auditory Primitives. *Current Directions in Psychological Science*, 9(2): 59–63.

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. https://doi.org/10.1016/S0010-0277(98)00069-9

Schyns, P. G., Petro, L. S., & Smith, M. L. (2009). Transmission of Facial Expressions of Emotion Co-Evolved with Their Efficient Decoding in the Brain : Behavioral and Brain Evidence. *PLoS ONE*, 4(5), e5625. https://doi.org/10.1371/journal.pone.0005625

Shapley, R., & Enroth-Cugell, C. (1984). Chapter 9 Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3, 263-346. https://doi.org/10.1016/0278-4327(84)90011-7

Smith, F. W., & Schyns, P. G. (2009). Smile Through Your Fear and Sadness : Transmitting and Identifying Facial Expression Signals Over a Range of Viewing Distances. *Psychological Science*, 20(10), 1202-1208. https://doi.org/10.1111/j.1467-9280.2009.02427.x

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and Decoding Facial Expressions. *Psychological Science*, 16(3), 184-189. https://doi.org/10.1111/j.0956-7976.2005.00801.x

Smith, M. L., & Merlusca, C. (2014). How task shapes the use of information during facial expression categorizations. *Emotion*, 14(3), 478-487. https://doi.org/10.1037/a0035588

Stein, T., Seymour, K., Hebart, M. N., & Sterzer, P. (2014). Rapid Fear Detection Relies on High Spatial Frequencies. *Psychological Science*, 25(2), 566-574. https://doi.org/10.1177/0956797613512509

Sweeny, T. D., Suzuki, S., Grabowecky, M., & Paller, K. A. (2013). Detecting and categorizing fleeting emotions in faces. *Emotion*, 13(1), 76-91. https://doi.org/10.1037/a0029193

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, 11(10), 697-709. https://doi.org/10.1038/nrn2889

Van der Schaaf, V. A., & van Hateren, J. V. (1996). Modelling the power spectra of natural images: statistics and information. *Vision research*, *36*(17), 2759-2770.

Vlamings, P. H. J. M., Goffaux, V., & Kemner, C. (2009). Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information? *Journal of Vision*, 9(5), 12-12. https://doi.org/10.1167/9.5.12

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. 2001 Effects of Attention and Emotion on Face *Processing in the Human Brain*: An Event-Related fMRI Study, 13.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, 6(6), 624-631. https://doi.org/10.1038/nn1057

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19(9), 1395-1407. https://doi.org/10.1016/j.neunet.2006.10.001

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PLOS ONE*, 12(5), e0177239. https://doi.org/10.1371/journal.pone.0177239

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of neuroscience*, *18*(1), 411-418.

Zhao, Y., Zhen, Z., Liu, X., Song, Y., & Liu, J. (2018). The neural network for face recognition : Insights from an fMRI study on developmental prosopagnosia. *NeuroImage*, 169, 151-161. https://doi.org/10.1016/j.neuroimage.2017.12.023

3.3 Transition

Dans l'Article 2, nous avons reproduit un paradigme de choix saccadique opposant un visage neutre à un visage émotionnel, joyeux ou apeuré. Cette fois, les images étaient présentées en HFS, en BFS ou non filtrées. Pour un groupe de participants, le contraste était égalisé entre les fréquences spatiales, tandis que pour un autre groupe de participants, le contraste n'était pas égalisé entre les fréquences spatiales. Les participants avaient pour instruction de faire une saccade vers le visage neutre dans une session, et vers le visage émotionnel dans une autre session. Les résultats ont montré que les saccades vers les visages émotionnels étaient effectuées plus rapidement que les saccades vers les visages neutres. Elles étaient aussi plus souvent dans la bonne direction, particulièrement lorsque le visage émotionnel était joyeux. Les saccades étaient plus souvent dans la bonne direction lorsque les images étaient présentées en HFS plutôt qu'en BFS (en particulier avec un visage émotionnel joyeux). Elles étaient aussi effectuées plus rapidement, un effet qui était seulement significatif lorsque le contraste était égalisé. L'analyse de saillance *bottom-up* effectuée sur les stimuli utilisés lors de l'expérience a mis en évidence les zones qui se distinguent par leurs caractéristiques physiques. Nous avons montré que la saillance de la région de la bouche (contrairement à celle des yeux) corrélait avec les performances des participants dans l'expérience comportementale.

Dans la suite de ce chapitre, nous présentons un modèle CNN qui nous permet de générer des cartes de saillance qui cette fois prennent en compte la tâche des participants, car le modèle apprend à discriminer les visages émotionnels et neutres. Ce modèle est présenté sous la forme d'un article indépendant. Similairement à ce qui a été fait dans l'Article 2, nous allons tester le lien entre les résultats des participants dans l'expérience comportementale et la saillance de la région de la bouche. Plus précisément, nous allons tester la capacité des cartes du CNN à prédire les résultats de l'expérience, et comparer les cartes de saillance issues du CNN aux cartes de saillance *bottom-up*. Ainsi, nous allons utiliser à nouveau les résultats de l'expérience comportementale de l'Article 2, bien que la manière d'analyser les données va différer sur certains points. Pour faciliter la compréhension de l'Article 3, l'expérience comportementale de l'Article 2 est réintroduite, et nous ne différencions plus les conditions de contraste. Dans l'Article 2, nous avons étudié la proportion de saccades correctes ainsi que les latences des saccades. Dans l'Article 3, nous avons étudié à la proportion de saccades correctes ainsi que les points d'arrivée des saccades.

3.4 Article 3

Dans ce troisième article, nous nous sommes intéressés aux régions utiles à la discrimination des visages émotionnels et neutres. Nous avons développé un CNN dont le but était de simuler l'expérience comportementale de l'Article 2. Pour rappel, dans cette expérience, un visage émotionnel était opposé à un visage neutre et les participants devaient faire une saccade le plus rapidement possible vers le visage émotionnel dans une session et vers le visage neutre dans une autre session. Les images étaient présentées

en BFS, en HFS ou non filtrées. Le contraste des images était égalisé pour la moitié des participants. Dans l'Article 2, nous avons utilisé des cartes de saillance *bottom-up* pour mettre en évidence les régions qui se distinguent d'un point de vue physique. Ces cartes font ressortir principalement 2 régions : les yeux et la bouche. Nous avons ensuite montré que la saillance de la bouche pouvait expliquer nos résultats comportementaux. Ici, l'objectif était d'aller plus loin, en utilisant des cartes de saillance générées à partir d'un CNN. Comme pour la simulation présentée dans le Chapitre 2, le CNN était entraîné à discriminer des paires de visages émotionnel-neutre et des paires de visages neutre-émotionnel (le visage émotionnel se trouvant soit à droite soit à gauche de la paire). Nous n'avons pas utilisé le même modèle que dans le Chapitre 2 (dans lequel un MLP était utilisé), car il était basé sur un résumé des images et non sur les images directement, ce qui rendait la visualisation des cartes de saillance impossible. Pour travailler directement sur des images, les CNN sont les modèles les plus adaptés. Les cartes de saillance générées à partir du CNN permettent de quantifier l'importance de chaque pixel pour réussir la tâche. Ainsi, en comparaison aux cartes *bottom-up*, ces nouvelles cartes permettent de prendre en compte la tâche des participants. Avec ce modèle, l'objectif était aussi d'analyser les performances du réseau dans chaque condition et de les comparer à celles des participants, similairement à ce qui avait été fait dans le Chapitre 2.

Guidés par les résultats de l'Article 2, nous nous attendions à ce que la région de la bouche soit particulièrement importante pour le réseau. De plus, nous nous attendions à ce que la saillance de la bouche, qu'elle soit générée à partir des cartes *bottom-up* ou des cartes du CNN, prédise significativement les performances des participants. Mais, nous nous attendions à ce que les prédictions soient meilleures avec les cartes du CNN. Tandis que les cartes de saillance *bottom-up* ont révélé plusieurs régions saillantes sur les visages (par exemple les yeux, la bouche, les cheveux...), les cartes de saillance basées sur le CNN ont révélé principalement une seule région saillante : la bouche. Ce qui suggère qu'elle a un rôle important pour la tâche, c'est-à-dire la discrimination de visages émotionnels et neutres. Des régressions linéaires simples ont montré que la saillance de la bouche, calculée à partir des cartes de saillance basées sur le CNN et des cartes de saillance *bottom-up*, prédisait significativement les performances des participants. Néanmoins, les prédictions basées sur le CNN étaient meilleures que celles basées sur la saillance *bottom-up*. En planifiant cette expérience, nous voulions aussi tester la capacité des performances du réseau à prédire celle des participants. Néanmoins, les performances du réseau étaient très bonnes et similaires dans toutes nos conditions expérimentales. Elles ne permettaient donc pas de prédire les performances des participants. Dans l'ensemble, nos résultats soulignent l'importance de la région de la bouche pour la détection des visages émotionnels et neutres, ainsi que l'utilité des cartes de saillance d'un CNN pour mettre en évidence des caractéristiques diagnostiques et prédire le comportement humain dans une tâche spécifique.

L'Article 3 a été soumis en mai 2021 dans la revue *Computational Intelligence and Neuroscience*. Le Tableau 3.2 dresse la liste des contributions de chaque auteur.

Contributeurs	Contributions
Léa Entzmann	Conception de la simulation ; Simulation ; Analyse des données ; Rédaction du manuscrit ; Édition du manuscrit
Nathalie Guyader	Conception de la simulation ; Édition du manuscrit
Martial Mermillod	Conception de la simulation ; Édition du manuscrit

Table 3.2 – Contributions des auteurs de l'Article 3.

Convolutional Neural Networks saliency map predicts fast detection of facial emotions in humans

Léa Entzmann[1, 2], **Nathalie Guyader**[2], **Martial Mermillod**[1]

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000, Grenoble, France.

[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France.

Keywords: Facial Expressions, CNN-based Saliency Maps, Bottom-up Saliency Maps, Spatial Frequencies, Saccadic Eye Movements

Abstract

In research on visual perception, it is argued that the low spatial frequencies, which carry coarse information, enable the rapid detection of threat-related stimuli, such as fearful faces. However, in a recent study using eye tracking we showed that the explicit detection of emotional faces was easier for high spatial frequencies, which carry finer details, especially with a happy emotional face (Entzmann et al., 2022). Moreover, the saliency of the mouth was found to correlate with participants performance. In the current study, we developed a convolutional neural network (CNN) to simulate this eye-tracking experiment, where participants had to discriminate an emotional from a neutral face. We analysed CNN-based saliency maps to identify the features that were used by the CNN. Contrary to classical bottom-up saliency maps, which are based on perceptual saliency only, such maps allow the evaluation of task-related saliency. Our results showed that, whereas bottom-up saliency maps revealed several salient regions, CNN-based saliency maps revealed principally one salient region: the mouth. This suggests

an important role of the mouth to succeed in the detection of emotional faces. Simple linear regressions showed that the mouth saliency, computed from both CNN-based and bottom-up saliency maps, was a significant predictor of participants' performance. Moreover, CNN-based predictions were better than predictions based on bottom-up saliency. Overall, our results underline (1) the importance of the mouth region for the detection of emotional and neutral faces, and (2) the usefulness of CNN-based saliency maps to highlight important features and predict human behaviour in a specific task.

1 Introduction

For humans, the ability to rapidly detect faces and their emotions is crucial, as they convey useful information for both survival and adapted social interactions. In neuroscience, it has been suggested that the human brain evolved to rapidly detect and process threatening stimuli, such as fearful faces (see for a review Tamietto and de Gelder, 2010). This fast detection is believed to emerge from a subcortical pathway that would only transmit coarse (i.e., low spatial frequency) information to the amygdala, a brain structure involved in evaluating emotional stimuli. This view is notably supported by recent intracranial electrophysiological data, showing an early amygdala activity differentiating fearful from happy and neutral faces (Méndez-Bértolo et al., 2016). In this study, faces were presented either in low spatial frequencies (LSF; frequencies carrying coarse information), high spatial frequencies (HSF; frequencies carrying detailed information), or broad spatial frequencies (BSF; unfiltered images). The early activity elicited by fearful faces was only observed in LSF or BSF, when the coarse information was available.

Although it is still under debates, several authors suggested that this fast and coarse detection of threatening stimuli is automatic, that is, mandatory and independent of task demand (e.g., Adolphs, 2008; Öhman, 2005). However, at a behavioural level, there is still no clear evidence that emotional faces capture attention in such an automatic way (see for a review Mulckhuyse, 2018). In recent eye-tracking experiments, our team studied the detection of emotional and neutral faces (Entzmann et al., 2021; Entzmann et al., 2022). We used a saccadic choice task, in which saccades (i.e., rapid eye move-

ments) are used instead of manual responses. Two faces, one emotional (happy or fearful), and one neutral, were displayed simultaneously on a screen and participants were asked to make a saccade toward a target face (the emotional or the neutral face). In Entzmann et al. (2021) only BSF images were used, and we observed that emotional faces were easier to detect (i.e., higher accuracy and/or shorter reaction time) than neutral faces, especially when the emotional face was happy. In Entzmann et al. (2022; an unpublished study available as a preprint) faces were presented in either LSF, HSF or BSF, and we observed that the accuracy for detecting emotional faces was higher when images were in HSF than LSF, especially when the emotional face was happy. This last result does not confirm the hypothesis that fearful faces are rapidly detected based on LSF information. Overall, we can suggest that, in this specific task, useful information was conveyed by HSF, and was more prominent with a happy face.

The main goal of the present study is to highlight the features that are diagnostic for the discrimination of an emotional and a neutral face, and to test whether this information is predictive of the outcome of a behavioural experiment. For that, we propose to use a convolutional neural network (CNN) for creating task-related saliency maps, and the data from our previous saccadic choice task for assessing their predictive ability (Entzmann et al., 2022).

In the literature, several studies have been interested in identifying the features that are useful for tasks related to the recognition of facial expressions. Some authors used eye-tracking recordings while participants were categorising the expression of a face (for example, happy, fearful, sad, angry, surprised, disgusted). Results from such experiments highlighted the importance of the eye and the mouth regions (Eisenbarth and Alpers, 2011) or the eye, nose, and mouth regions (Schurgin et al., 2014). They also revealed different patterns depending on the expression being scanned. For example, the reliance on the mouth was higher for happy faces, whereas for fearful faces the reliance on the eyes and the mouth was more balanced (Eisenbarth and Alpers, 2011; Schurgin et al., 2014). In our behavioural experiment, eye movements were used as participants' answer, to access the speed of visual processing but not to highlight which parts of the face attracted the gaze. Indeed, the participants made their decision with their gaze in the centre of the screen, and then initiated a saccade toward the target, which gen-

erally landed around the nose region. Then, we were unable to access eye fixations while participants were performing the task. Other studies used techniques consisting of masking some parts of the faces to isolate the areas that were the most relied on to categorise expressions (Gosselin and Schyns, 2001; F. W. Smith and Schyns, 2009; M. L. Smith et al., 2005; Wegrzyn et al., 2017). For example, several studies used a Bubbles methodology, which allows the identification of task-relevant features by comparing categorisation performance when different image parts are masked. Results from such studies showed that the mouth was particularly important for the discrimination between happy and neutral faces (Gosselin and Schyns, 2001), and that the mouth was important to recognise happiness whereas the eyes were important to recognise fear (F. W. Smith and Schyns, 2009; M. L. Smith et al., 2005).

In deep learning, especially with CNN, task-related saliency computation methods have emerged to highlight features relevant for the prediction of a trained model (e.g., Simonyan et al., 2014). Generally, task-related saliency computation methods are used as a visualisation tool that follows the introduction of a model whose purpose is not the comparison with humans. For example, several papers highlighted the face parts that were used by their models in a facial emotion recognition task. Minaee and Abdolrashidi (2021) observed that the mouth was important for their CNN to recognise fear and happiness, whereas a wider region was required for neutral face recognition. For two other models, authors found that the mouth was important for recognising happiness, and the eyes were important for recognising fear (Dailey et al., 2002; Jiao et al., 2019). Also, a very large number of CNN models were proposed to predict visual attention in humans (Huang et al., 2019; Kruthiventi et al., 2016; Kümmerer et al., 2014; Kümmerer et al., 2016; Pan et al., 2016; Vig et al., 2014). Generally, the goal of those models is to highlight the region in which the observer's eyes may focus first, in a task-free visual exploration. Eye-tracking benchmarks are used to associate natural scene images with recorded human eye fixations in a task-free visual exploration. But more recently some models incorporate task-related saliency using CNN parameters (Mahdi et al., 2019; Murabito et al., 2018) or task-related eye-tracking benchmarks (Zheng et al., 2018) to predict visual attention in a specific task. Their results underlined the importance of considering not only bottom-up but also task-related saliency to predict

visual attention in a specific task, and the potential of CNN to assess it.

In this study, we computed a simple CNN model to simulate our previous behavioural experiment (Entzmann et al., 2022) in which participants have to make a saccade toward a target face (the emotional or the neutral one). To correctly perform the task, participants have to discriminate the emotional and the neutral face. Hence, the network was presented with pairs of faces from the experiment and was trained to categorise them depending on whether the emotional face was on the left and the neutral face on the right side of the pair (Emotional/Neutral class) or the opposite (Neutral/Emotional class).

First, we computed the accuracy of the CNN in each experimental condition to test the hypothesis that similar recognition patterns can be obtained with both humans and a statistical tool. Indeed, the use of an artificial neural network allows the comparison of the performance of humans to that of an emotionless algorithm. In the behavioural experiment, we can suppose that there are more statistical differences between the two faces in HSF, particularly with a happy face (or that the useful cues are more prominent in this condition). This would explain why LSF fearful faces did not attract gaze more than other faces. Artificial neural networks are particularly useful to disentangle perceptual from emotional factors. In the 2000s, several studies used linear or multi-layer perceptrons to explain the emotional processing of faces by humans (Dailey et al., 2002; R. Li and Cottrell, 2012; Mermillod et al., 2010; Mermillod et al., 2009). A similarity between human and neural network emotion recognition accuracy has been directly highlighted by Dailey et al. (2002). The authors used a model called EMPATH based on a simple linear perceptron that takes as input a decomposition of images through Gabor filters. For both humans and the network, fear was the hardest emotion to recognise, and happiness was the easiest. The authors explain their results assuming that the recognition of emotions can rely only on the statistical properties of images. In such studies, emotion recognition was assessed by categorising a single face, the, results are still not directly comparable to those of the saccadic choice task. Deep neural networks, especially CNN are currently the models that achieve the best performance in facial emotion recognition (Huang et al., 2019; Kartali et al., 2018; S. Li and Deng, 2020; Mollahosseini et al., 2016; Rouast et al., 2019; Vyas et al., 2019). In the domain

of facial emotion recognition, similarities can be found while comparing categorisation performance for different emotions across different studies. For example, for both humans (Calvo and Nummenmaa, 2016; Tottenham et al., 2009) and CNN (Khanzada et al., 2020; Mollahosseini et al., 2016; Pitaloka et al., 2017), happiness is often the easiest emotion to recognise, whereas fear is often the hardest. However, to our knowledge, no study directly compared to recognition of facial emotions in humans and CNN.

Then, we computed CNN-based saliency maps to see which pixels were important for the task. We compared these maps with classical bottom-up (i.e., image-related saliency) saliency maps, which highlight the salient regions of the faces based on their physical characteristics. Bottom-up saliency can be computed based on local differences in intensity and orientation on a specific image, and highlights the pixels that stand out the most from their neighbours (see for a review Borji and Itti, 2012). In our previous study, we analysed the relationship between the bottom-up saliency of the different features and the performance of participants. Bottom-up saliency maps highlighted different parts of the faces: the eyes, the mouth and facial contours. We showed that the performance of the participants correlated with the saliency of the mouth region, and not with that of the eye region. This suggests that the mouth region may be more useful to discriminate the emotional and the neutral face. Then we expected the CNN-based saliency maps to particularly emphasise the mouth area.

Finally, we assessed the ability of CNN-based saliency maps to predict human behaviour in the saccadic choice task (human accuracy, saccade endpoints), and compare it with that of bottom-up saliency maps. Indeed, in our previous study we have already shown that the mean bottom-up mouth saliency in an experimental condition (i.e., with an emotional or a neutral target, in BSF, HSF, or LSF, with a happy, or fearful emotional face) was linked with the participants' performance in this condition (Entzmann et al., 2022). Here, we suppose both bottom-up and task-related mouth saliency can be significant predictors of the performance of participants in one experimental condition. But we supposed that the incorporation of task-related saliency would lead to better results. In fact, the importance of both bottom-up and task-related saliency in visual attention in humans is largely documented in the literature (e.g., Ballard and Hayhoe, 2009; Murabito et al., 2018; Noudoost et al., 2010; Zelinsky et al., 2005). And, CNN-

based saliency maps, although they are task-related, may in addition reflect bottom-up differences in a way that is difficult to quantify (Adebayo et al., 2018). We also tested whether the mouth saliency computed from both bottom-up and CNN-based saliency maps can predict the vertical position of saccades endpoints. The assumptions were that the more salient the mouth is in one experimental condition, the lower (i.e., closer to the mouth) the endpoints will be, and the higher the human accuracy will be.

Overall, it is important to specify that the model that we present here is not new, and quite simple as we do not wish to achieve the best accuracy and used very stereotypical images. The CNN architecture was adapted from the Lenet5 model (Lecun et al., 1998), and associated with a gradient-based visualisation technique (Simonyan et al., 2014). The originality of this study is to propose an application of such computational methods for the understanding of human behaviour. Thus, this study links techniques from the field of computer science with questions from the field of neuroscience and cognitive psychology.

2 Method

2.1 Dataset

The dataset consisted of the exact photographs of emotional and neutral faces used in Entzmann et al. (2022). There were 20 different individuals (10 men, 10 women), chosen among the Karolinska Directed Emotional Faces database (Lundqvist et al., 1998). Each face portrayed either a neutral, happy or fearful expression, and was presented in different spatial frequencies, i.e., BSF, HSF and LSF. In the behavioural experiment, participants were divided into two groups, one in which the luminance contrast was equalized between images (particularly between LSF and HSF), and one in which the contrast was not equalized. For simplification and since luminance contrast, equalized or not, had no impact neither on human or CNN accuracy, results are presented regardless of contrast conditions. An example of a face within different spatial frequencies is given in Figure 1 (a). Images were then gathered in pairs with a male and a female, and an emotional and a neutral face. In each pair, both faces had the same spatial fre-

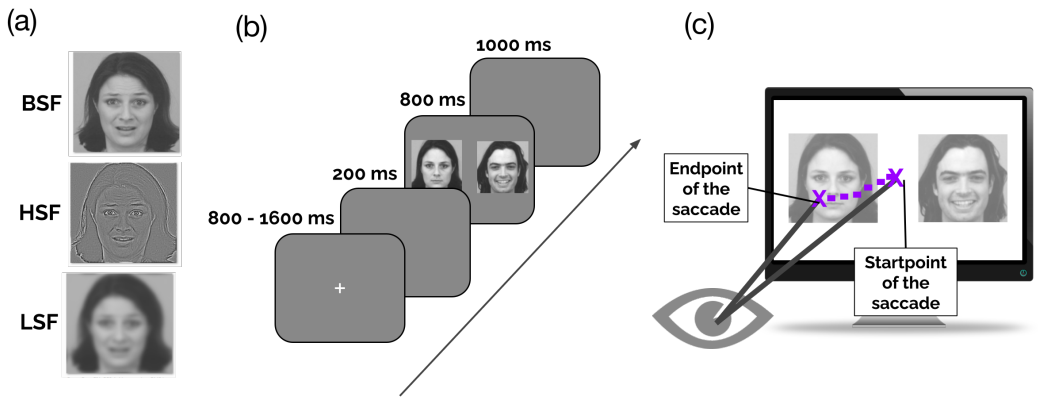

Figure 1: (a) Example of a face in the different spatial frequency conditions: BSF, HSF and LSF. (b) Time course of one trial of the behavioural experiment. (c) Visualisation of a saccadic response.

quency and luminance-contrast condition (BSF, LSF, HSF, and for LSF and HSF with or without contrast equalization). Overall, this led to 4000 different pairs of faces.

2.2 Behavioural experiment

In this part, we briefly describe the experiment in Entzmann et al. (2022).

Participants: Seventy-eight human participants were included in the analysis (39 females; mean age ± SD: 21.39 ± 0.98 years; age range: 18–33 years). Participants gave their informed written consent before the experiment.

Material: Stimuli sized 300x300 pixels and were displayed on a 24-inch screen with a spatial resolution of 1360×768 pixels. In this configuration, each image sized 11x11° of visual angle. An Eyelink 1000 Plus (SR Research) eye-tracker was used to record eye movements.

Procedure: Participants performed two sessions of 240 trials; in one session the target was the emotional face and in the other, it was the neutral face. Each session lasted approximately 15 minutes, and their order was counterbalanced between participants. Sessions were divided into three blocks: one with BSF pairs, one with HSF pairs, and one with LSF pairs. During the experiment participants had their head stabilized at a

fixed distance of 57 cm away from the screen. At each trial, participants first saw a cross in the centre of the screen, that they had to fixate during a pseudo-random time interval ranging between 800 and 1600 ms. After a 200 ms gap, a neutral and an emotional face (happy or fearful) were displayed during 800 ms, one of each side of the screen. Participants were asked to make a saccade as quickly as possible toward the target (the emotional or the neutral face). Only the first saccade was analysed. The saccade was considered as correct if it was directed toward the target, and, as incorrect otherwise. The centre of each image was located at a distance of 8° from the centre of the screen. Images were then followed by a gray screen lasting 1000 ms. A visual representation of a trial (b) and a saccadic response (c) are displayed in Figure 1. The proportion of correct saccades was used as a measure of participants' performance.

Saccade endpoints: Although not presented in the original study, the analysis of saccade endpoints is presented in this paper. In fact, they can reflect which parts of the face attract the attention, and may be in relation with the mouth saliency. Similarly to a previous study (Entzmann et al., 2021), we extracted the Y-coordinate Y_e of the first saccade endpoint. This coordinate was calculated in the image space, thus, within a square of $11 \times 11°$ of visual angle, the coordinates X_0 and Y_0 being at the top-left corner. Then, we analysed the vertical distance between between Y_e and the Y-coordinate of the centre of the image Y_c (see Figure 2). Our dataset was created in such a way that the eyes corresponds to the centre of the image. Therefore, this measure can be seen as the vertical distance between the saccade endpoint and the eyes; a negative value meant that the endpoint was below the eyes.

2.3 Model Specification

We built a CNN to categorise pairs based on the position of the emotional and the neutral face. The network had to decide whether an input pair was an Emotional/Neutral or Neutral/Emotional pair. The architecture of the CNN was adapted from the LeNet5 model presented in Lecun et al. (1998). We choose this model for its simplicity, as our task is quite simple (stereotypical faces, front view), we did not move to higher complexity. As for every neural network, the goal of the model was to represent an arbitrary

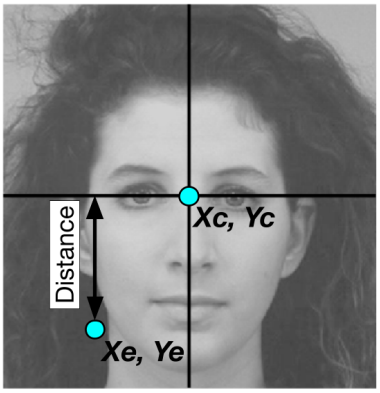

Figure 2: Visualisation of a saccade endpoint, defined by its coordinates X_e and Y_e, and of the measure of the vertical distance to the centre. This measure corresponds to the distance between between Y_e and the Y-coordinate of the centre of the image Y_c.

function $y = f(x)$, using a succession of one input layer, a number of hidden layers and one output layer. In our case, y is the realisation of the variable Y, representing 2 classes, and x is a realisation of X, representing an image with a resolution of $m \times d$ (300x600 pixels). As describe by Jospin et al. (2020), in the simplest architecture of feedforward networks, each layer $h \in [1, H]$ is represented as a linear transformation, followed by a nonlinear operation s, also known as activation function. Then, the architecture of an artificial neural network can be formulated as follows:

$$\begin{aligned} l_0 &= x \\ l_h &= s_h(w_h l_{h-1} + b_1) \\ y &= l_H \end{aligned} \tag{1}$$

where l_h represents the information obtained after the h-th layer of the network, and H the number of layer in the network. Such an architecture depicts a set of functions isomorphic to the set of possible parameters $\Theta = (w, b)$ where w are the weights of the network and b the biases. During a training process, regressions are applied to the parameters Θ based on the training data, an association of inputs x and their corre-

sponding labels y. For our model, the input size was 300x600, which is higher than that images from classical datasets used in machine learning, but corresponds to the size of the pairs of images that were presented to participants. The CNN had to categorise pairs based on the position of the emotional and the neutral face. Hence, the network had to decide whether the pair was an Emotional/Neutral pair or a Neutral/Emotional pair. This task is similar to the behavioural task, as participants have to discriminate the two faces to detect in which side the emotional or the neutral face is.

The implementation was built in Python using Keras, with Tensorflow as a backend. Only a few changes were made from the initial architecture comprising 7 layers (not counting the input). Therefore, as in the initial architecture, the network was first composed of 2 convolutional layers, each of which is followed by a pooling layer. Then, the network was composed of 2 fully-connected layers, followed by the output layer. A visual representation of the proposed CNN model is given in Figure 3, and each type of layer is more fully describe below:

Convolutional layers can be seen as feature extractors; each output unit is connected to a small neighborhood in the input. Here, the first convolutional layer is composed of 6 filters (kernel of size 5x5 pixels). The second convolutional layer is composed of 16 filters (kernel of size 5x5 pixels). After each convolutional layer, the information is transformed using a ReLu activation function, and transmitted to a pooling layer.

Pooling layers are used to reduce the size of the feature maps after each convolutional layer. Here, size reduction was performed by taking the maximum value of a map over a window of size 2 (i.e., a Max Pooling operation was performed). After the last pooling layer, the information is flattened and transmitted to fully-connected layers.

Fully-connected layers are used after the convolutional and pooling layers to perform the classification. The first fully-connected layer is composed of 120 units, and the second fully-connected layer is composed of 84 units. The information is then transmitted to the output layer, after which a Softmax activation function is used, generating outputs between 0 and 1. The output is composed of 2 units corresponding to the probability of the Emotional/Neutral class, and the probability of the Neutral/Emotional class. Finally, the predicted class was calculated based on the highest probability of the output vector.

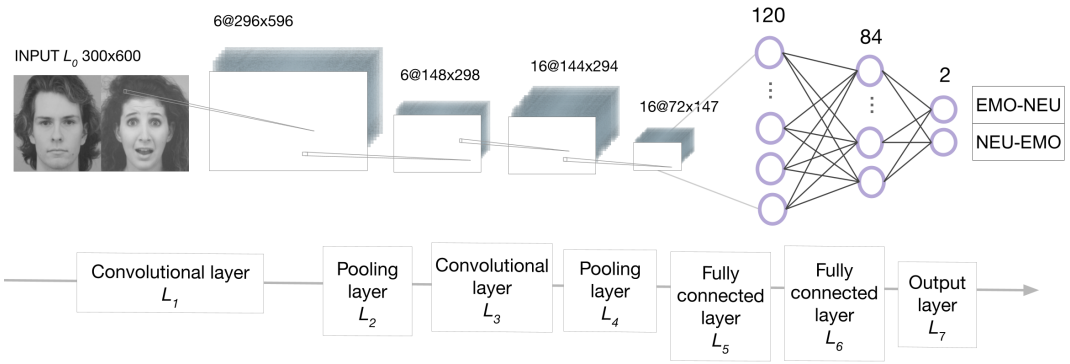

Figure 3: Visual representation of the CNN. The input is a pair of images with a size of 600x300 pixels. The output is a vector of size 2, the first value corresponds to the probability of being in the Emotional/Neutral class, and the second output to the probability of being in the Neutral/Emotional class.

2.4 Simulations

The simulation was composed of 200 runs (i.e., 200 different training and test procedures with the same network architecture), to ensure a stable and reliable analysis. A run began with the distribution of pairs between the training, the validation, and the test set. The validation set was used to test the generalisation performance of the network during the training. Note that we control that each individual was only in one set. More precisely, 6 men and 6 women were randomly chosen for the training set, 3 men and 3 women for the test set and the remaining for the validation set. This led to 1400 pairs in the training set, 40 pairs in the validation set, and 216 pairs in the test set. During the training, each pair was associated with a vector of size 2 coding for its category, Neutral/Emotional ([0,1]) or Emotional/Neutral ([1,0]), depending on whether the emotional or the neutral face is on the left or on the right side. Pairs were transmitted to the network within a batch size of 32 (for the training and test set only), and associated with their outputs over 30 epochs using a standard back-propagation algorithm. An early stopping was scheduled when the accuracy on the validation set showed no improvement for 5 epochs. The classification error was evaluated using a categorical

cross-entropy (CCE) loss, to be minimized during the training:

$$CCE = -\sum_{b=1}^{B} y_b.log(\hat{y}_b) \tag{2}$$

Considering that categories are encoded with a one-hot encoding (i.e., the labels code for 2 classes on $B = 2$ output units, with only one which takes the value 1), $\hat{y}_b$ is the model prediction on the b-th output unit, and y_b is the correct value. The weight update was performed using the Adam optimization algorithm with a learning rate of 0.001 (Kingma and Ba, 2017).

After the training, the network was tested on its ability to dissociate Emotional/Neutral from Neutral/Emotional pairs on the test set. An accuracy score is defined as follows:

$$ACC = \frac{1}{N}\sum_{n=1}^{N} 1_{\{y_n=\hat{y}_n\}} \tag{3}$$

where N is the sample size of the test set, and y_n the true class, and $\hat{y}_n$ the predicted class, $\forall n \in \{1, \cdots, N\}$. Thus, each pair is associated with a value depending on whether the categorisation is correct (1) or not (0), and the accuracy score represents the mean of those values. The closer it is to 1, the higher the classification performance of the CNN is.

2.5 CNN-based saliency

To visualise which pixels of the image are important to discriminate the neutral from the emotional face, we computed CNN-based image-specific class saliency maps (more simply referred here as CNN-based saliency maps) for each pair of the test dataset using the gradient-based method (Simonyan et al., 2014). In the context of a CNN, the idea is to approach the equation (1) with a linear function in the neighbourhood of x_0 by computing the first-order Taylor expansion:

$$f(x) \approx w^T x + b \tag{4}$$

where w can be seen as the derivative of f with respect to the image x at the point (image) x_1:

$$w = \frac{\partial f}{\partial x} \mid_{x_1} \tag{5}$$

Then, given the input image x_1, each pixel of the class saliency map in the i-th row and j-th column is computed as:

$$M_{i,j} = |w_{h(i,j)}| \tag{6}$$

Where w is the gradient vector of the class score with respect to the input image (as we used gray scale image, the number of elements in w corresponds to the number of pixel in x_1) and $h(i, j)$ is the index of the element of w, corresponding to the image pixel in the i-th row and j-th column. Therefore, for a given image and a given class, the gradient of the loss function is backpropagated to the input layer, and displayed in the image space. This method allows finding the pixels for which a change would affect the output the most, and can also be used for object localisation or segmentation, as important image pixels usually correspond to the object to be detected. With this procedure, we obtained CNN-based saliency maps for each image of the test set. We then computed mean maps for each experimental condition (in HSF, BSF or LSF, with a happy or fearful emotional face). More precisely, each pixel of a mean saliency map corresponded to the mean values of this pixel taken from all saliency maps belonging to the same experimental condition. This is appropriate here because each face is centered at the same position in the image. Finally, we also normalized mean saliency maps in order to obtain of minimum value of 0 and a maximum value of 255 (for pixel values ranging between and 255) for each map.

Using the CNN-based saliency maps, we obtained a value of importance for each pixel of the face. Guided by the results of our previous study, we were particularly interested in the importance of the mouth region, that we quantified for the different experimental conditions (for the emotional and the neutral face, for the different spatial frequencies, and when the emotional face is happy or fearful). We computed a score, the relative mouth saliency (RMS), which is defined as a weighted mean of the pixel of M in a specific selected mouth area A associated to the corresponding face area B (see Figure 4 (a) for a visual representation of A and B):

$$RMS_{CNN} = \frac{\sum_{(i,j)\in A} M_{i,j}}{\sum_{(i',j')\in B} M_{i',j'}} \tag{7}$$

This score was computed twice for each maps: once for the mouth of the emotional face, and once the mouth of the neutral face. The mouth area A was limited to

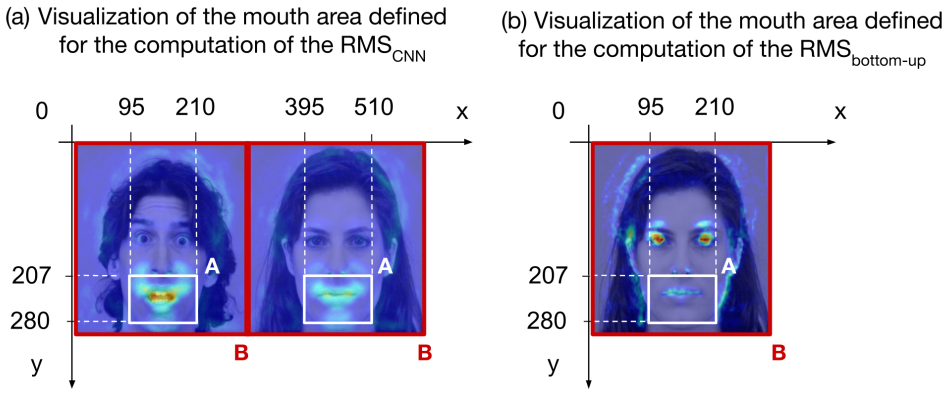

Figure 4: Visualisation of the mouth area A considered for the computation of (a) the RMS_{CNN} and (b) the $RMS_{Bottom-up}$. Scores are computed for each face area B. For the RMS_{CNN} the mouth area was limited to pixels with an X- coordinate between 95 and 210 (for the mouth of the face on the left side of the pair) or between 395 and 510 (for the mouth of the face on the right side of the pair), and a Y-coordinate between 207 and 280. For the $RMS_{Bottom-up}$ there was only one mouth area, corresponding to an X-coordinate between 95 and 210. For illustration purpose, maps are displayed on top of one randomly chosen pair from the dataset.

pixels chosen to contain the mouth of every face, irrespective of the expression or the individual. Overall, the RMS gives the proportion of importance contained in a mouth area.

2.6 Bottom-up saliency

As a comparison to CNN-based saliency maps, we computed maps from a classical bottom-up saliency model. More precisely, we used the computational saliency model proposed by Walther and Koch (2006), itself based on the Itti et al. (1998) model. We chose this model because it constitutes one of the main standard in psychology and cognitive neuroscience to compute saliency maps. The algorithm is inspired by the biology of the human visual system and considers intensities and orientations to attribute a saliency value to each pixel of an input image on the basis of pure bottom-up pro-

cesses (irrespective to the task-demand). The model was implemented in Matlab using the SaliencyToolbox (http://www.saliencytoolbox.net) and the makeSaliencyMap function. Note that, similarly to what have been done in our paper (Entzmann et al., 2022), we have changed the default settings for calculating the maps. At the end of this process, a saliency map was obtained for each image. Then, we also quantified the relative mouth saliency using the same procedure than for the calculation of the RMS_{CNN}. The RMS computed from the bottom-up saliency maps will be referred as $\text{RMS}_{Bottom-up}$. As in this case maps were computed on a single image, there was only one mouth area (see Figure 4 (b) for a visual representation).

2.7 Statistical analysis

To characterise the differences between the experimental conditions for the mean human accuracy, saccade endpoints of correct saccades, CNN accuracy, RMS_{CNN} and $\text{RMS}_{Bottom-up}$, statistical analysis were carried out. For the human accuracy and the saccade endpoints, analysis were performed on the mean values obtained for each participants in each experimental condition. For the CNN accuracy and the RMS_{CNN}, they were performed on the mean values obtained for each run in each experimental condition, and for the $\text{RMS}_{bottom-up}$, they were performed on the values obtained for each image. For the human accuracy, the saccade endpoints, and the RMS_{CNN}, a paired samples t-test with the Target (Emotional, Neutral) as a within-subject factor was used to assess the main effect of the target. Next, a repeated measures ANOVA with the Emotional Facial target (EFE; Happy or Fearful) and the Spatial Frequency (BSF, HSF, LSF) as within-subject factor was conducted when the target was the emotional face. Similarly, a repeated measures ANOVA with the Emotional Facial distractor (EFE; Happy or Fearful) and the Spatial Frequency (BSF, HSF, LSF) as within-subject factor was conducted when the target was the neutral face. For the $\text{RMS}_{bottom-up}$, similar analysis were carried out, the only difference is that the independant variables (the EFE, the Spatial Frequency and the Target) were coded as between (and not within) factors. For the network accuracy, as there was no target, a repeated measures ANOVA with the Emotional Facial Expression (EFE; Happy or Fearful) and the Spatial Frequency (BSF,

HSF, LSF) as within-subject factors was conducted. Finally, a Welch Two Sample t-test was perfomed to compare the $\text{RMS}_{bottom-up}$ and the RMS_{CNN}. Paired samples t-tests with Bonferroni corrections were used for pairwise comparisons, effect sizes were estimated by calculating partial eta-squared (η_p^2) for ANOVAs and Cohen's d for t-tests. An effect was considered significant if its p value was below the threshold $alpha$ = .05. In the results section only a summary of the results of the statistical analysis is reported, but the complete analysis are given in the Appendix.

3 Results

3.1 Human performance

Figure 5 (a) shows mean proportion of correct saccades observed in the behavioural experiment (Entzmann et al., 2022) when the target is an emotional (left column) or a neutral face (right column), in each spatial frequency condition, with a happy or a fearful emotional face. On average, the task was quite difficult, and the saccadic response was correct for 62,5% of the trials. Results revealed that the accuracy was overall higher when the target was emotional than neutral. When the target was the emotional face, the accuracy was higher with a happy than a fearful emotional face, and for images in BSF and HSF than LSF. Overall, with these behavioural results, presented in detail in the Appendix, we notably observed that emotional faces are more attractive than neutral faces and that their detection is easier for HSF than LSF pairs, and with happy emotional faces.

3.2 Saccade endpoints

Figure 5 (b) shows mean vertical distance to the centre (in degrees of visual angle) observed in the behavioural experiment (Entzmann et al., 2022) when the target is an emotional or a neutral face, in each spatial frequency condition, with a happy or a fearful emotional face. Results revealed that the endpoints went lower when the target was emotional than neutral. When the target was the emotional face, the endpoints were lower when the target was happy than fearful, and for images in LSF than HSF

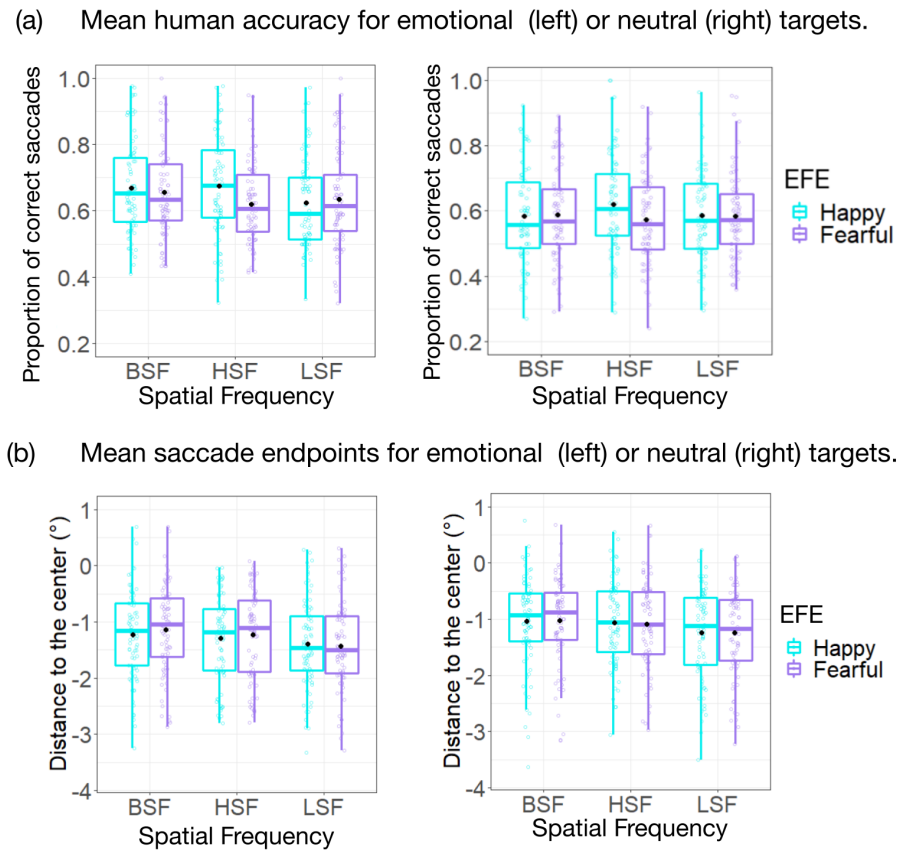

Figure 5: (a) Mean proportion of correct saccades, and (b) mean distance to the centre of correct saccades, for emotional (left column) and neutral (right column) targets, in the different spatial frequencies (BSF, HSFand LSF) and emotional facial expressions (EFE; happy or fearful).

or BSF. For a visual representation, Figure 7 (a) displays heat maps computed from saccade endpoints convoluted with a small 2D gaussian, in each condition and for all participants.

3.3 Model performance

Figure 6 shows mean accuracy in each spatial frequency condition, with a happy or a fearful emotional face. The CNN reached an average ACC of 0.957, corresponding to a correct classification for 97.5% of the tested pairs. This accuracy was signing a ceiling effect. Indeed, this performance was similar for each spatial frequency, whether the emotional face was happy or fearful.

3.4 CNN-based saliency maps

Mean CNN-based saliency maps in the different spatial frequencies, with a happy or fearful emotional face, are presented in Figure 7 (b) for the Emotional/Neutral class (saliency maps for the Neutral/Emotional class are not displayed but are are symmetric to that of the Neutral/Emotional class). Mean CNN-based saliency maps revealed that the mouth was critical for correct classification, in each condition. It was more informative than any other face parts. The nasolabial fold was also important. Moreover, we can see that the mouth of the emotional faces was more emphasised than the mouth of the neutral faces, meaning that the mouth of emotional faces was more informative to correctly perform the task. This suggests that, to discriminate an emotional from a neutral face, the network mainly relied on the mouth area, and especially on the cues of the emotional one. For comparison, bottom-up saliency maps are displayed in Figure 7 (c) in the same experimental conditions. We can see that they reveal a high saliency not only for the mouth region, but also for the eyes and facial contours. It is interesting to note that whereas the eyes benefit of a high bottom-up saliency (as revealed by bottom-up saliency maps) there were not used by the CNN to perform the task.

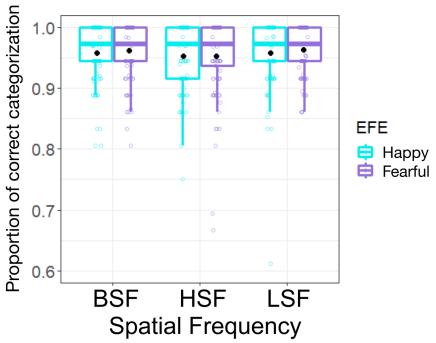

Figure 6: Mean CNN accuracy in the different spatial frequencies (BSF, HSF and LSF) and emotional facial expressions (EFE; happy or fearful). Note that their is no target condition here. This is because, contrary to the behavioural experiment, the network had to discriminate the two faces without targeting one face in particular.

3.5 RMS$_{CNN}$

Figure 7 (d) shows mean RMS$_{CNN}$ of emotional and neutral faces, in each spatial frequencies, with a happy or a fearful emotional face. The RMS$_{CNN}$ was overall higher for emotional than neutral faces. For emotional faces, The RMS$_{CNN}$ was overall higher for happy than fearful faces, and for images in BSF and HSF than LSF, as well as for images in BSF than HSF. With a happy face, the RMS$_{CNN}$ was higher in BSF than HSF, whereas this was the opposite with a fearful face. For neutral faces, the RMS$_{CNN}$ was higher for images in BSF than HSF and LSF.

3.6 RMS$_{Bottom-up}$

Compared to the RMS$_{CNN}$, the RMS$_{Bottom-up}$ was overall lower, with a mean value of 0.075 (the mean value for the RMS$_{CNN}$ being 0.13). Figure 7 (e) shows mean RMS$_{Bottom-up}$ of emotional and neutral faces, in each spatial frequencies, with a happy or a fearful emotional face. Note that in this analysis the RMS$_{Bottom-up}$ of neutral faces was duplicated to consider happy or fearful distractors. The purpose was to keep an homogeneity in the analysis, as for the human accuracy, the saccade endpoints and the

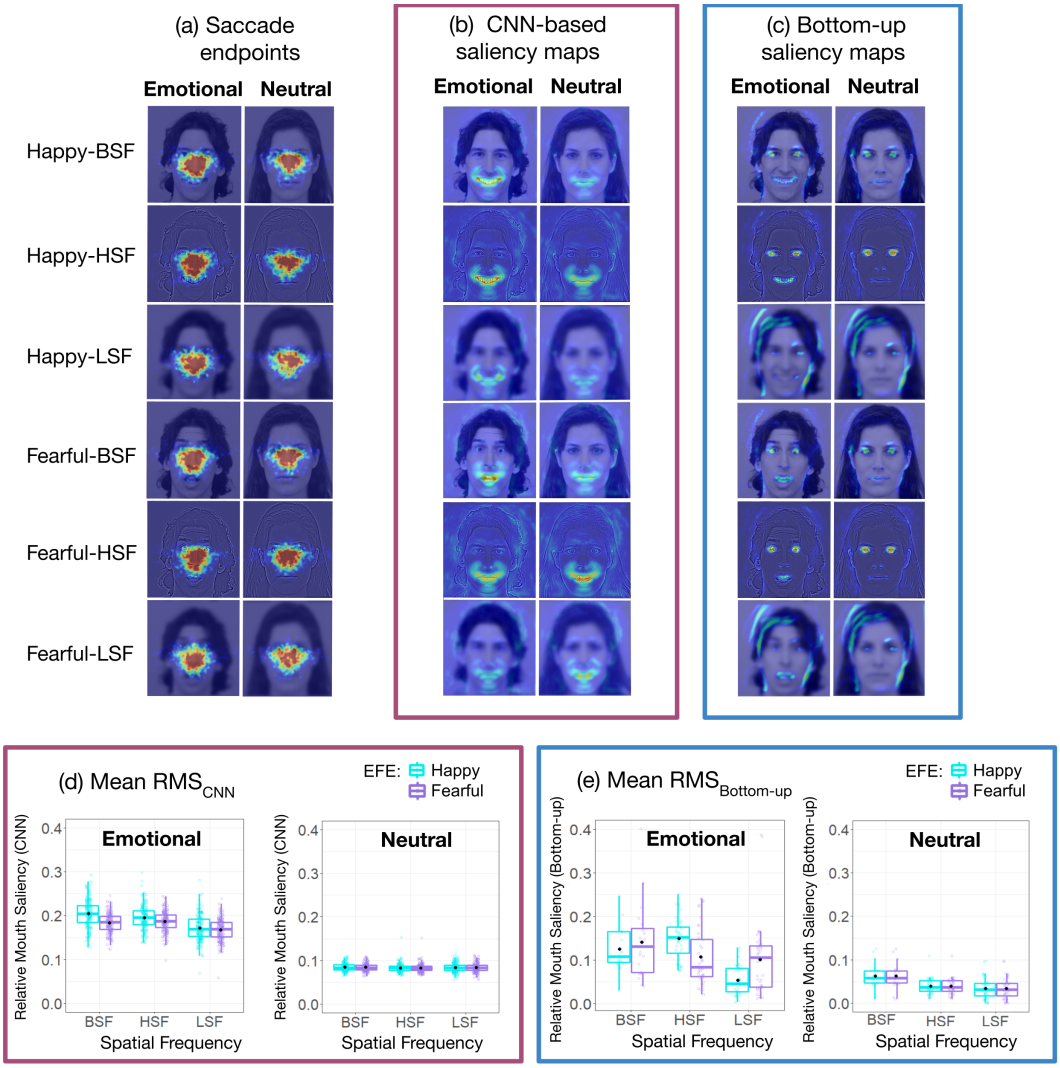

Figure 7: (a) Heat maps computed from all saccade endpoints of correct saccades, (b) mean CNN-based class saliency maps for the Emotional/Neutral class, (c) mean bottom-up saliency maps, (d) mean RMS_{CNN}, and (e) mean $RMS_{Bottom-up}$ in the different experimental conditions (i.e., for emotional and neutral targets, in each spatial frequencies, with a happy or fearful emotional face). For illustration purpose, maps are displayed on top of one randomly chosen pair from the dataset. Also, for the bottom-up saliency, the maps presented for neutral faces in one spatial frequency condition are the same whether the emotional face is happy or fearful.

RMS$_{CNN}$ this distinction was relevant. The RMS$_{Bottom-up}$ was overall higher for emotional than neutral faces. For emotional faces, the RMS$_{Bottom-up}$ was higher for images in BSF and HSF than LSF. In HSF, the RMS$_{Bottom-up}$ was higher for happy than fearful faces, whereas this was the opposite in LSF. Also, the higher saliency in HSF and BSF than LSF was only significant with a happy face. For neutral faces, the RMS$_{Bottom-up}$ was higher for images in BSF than in HSF or LSF.

3.7 Prediction of human behaviour on the basis of CNN saliency maps

Saccade accuracy: Here, we tested whether the saliency maps provided by the CNN can predict human performance in each condition thought the measure of the RMS$_{CNN}$. We computed the mean RMS$_{CNN}$ and the mean proportion of correct saccades (ACC_{human}) for different subset of data. Each subset corresponds to the experimental conditions of the behavioural experiment, i.e., in the different spatial frequencies, with a happy or fearful face, when the target is the emotional or the neutral face. Then, a simple linear regression was proposed to predict the proportion of correct saccades in each condition (ACC_{human}) based on the RMS$_{CNN}$:

$$ACC_{human} = \beta_0 + RMS_{CNN}\beta_1 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \tag{8}$$

A significant regression equation was found ($F(1,10) = 22.3$, $p < .001$), with an adjusted R^2 of 0.76. The estimated β_0 was 0.54 and the estimated β_1 was 0.58. Therefore, a high RMS$_{CNN}$ in one condition is linked to a high accuracy in this condition (Figure 8 (a)).

Saccade endpoints: In the same way we tested whether we can predict the vertical distance to the centre of saccade endpoints in each condition using the RMS$_{CNN}$. The goal was to test the hypothesis that the higher the RMS$_{CNN}$ is in one experimental condition, the lower (i.e. closer to the mouth) the endpoint will be. A significant regression equation was found ($F(1,10) = 5.3$, $p = .044$), with an adjusted R^2 of 0.28. The estimated β_0 was -1 and the estimated β_1 was -1.5. Therefore, a high RMS$_{CNN}$ in one condition is linked to a low endpoint in this condition (Figure 8 (b)).

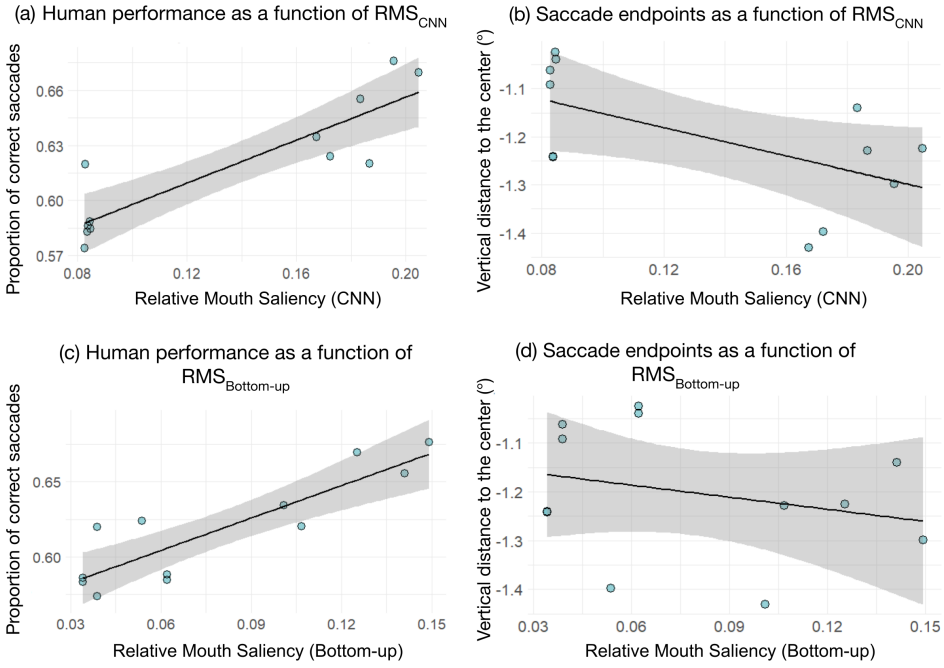

Figure 8: Scatter plot and simple linear regression model for the prediction of the proportion of correct saccade (left) or saccade endpoints (right) based on the RMS$_{CNN}$ (top) or based on the RMS$_{Bottom-up}$ (bottom). Each dot represents the mean proportion of correct saccades or the mean vertical distance to the centre over all participants combined with the mean RMS$_{CNN}$ over all runs of the CNN or the mean RMS$_{Bottom-up}$ over all images for one experimental condition (i.e., one dot corresponds to one target, EFE, and spatial frequency condition of the Figure 5).

3.8 Prediction of human behaviour on the basis of bottom-up saliency maps

Saccade accuracy: Here we tested whether we can predict the human performance in each condition using the RMS$_{Bottom-up}$. A significant regression equation was found ($F(1,10) = 32.3$, $p < .001$), with an adjusted R^2 of 0.74. The estimated β_0 was 0.56 and the estimated β_1 was 0.71. Therefore, a high RMS$_{Bottom-up}$ in one experimental condition is linked to a high accuracy in this condition (Figure 8 (c)).

Saccade endpoints: We tested whether we can predict the vertical distance to the

centre of saccade endpoints in each condition using the $\text{RMS}_{Bottom-up}$. No significant regression equation was found (Figure 8 (d)).

4 Discussion and conclusion

In conclusion, CNN-based saliency maps offered important insights and a better understanding of human performance on several points. First, they revealed that the discrimination of an emotional and a neutral face mainly relies on the mouth area. Moreover, the RMS_{CNN} computed from the CNN-based saliency was found to be a significant predictor of the accuracy of participants. The linear regression showed that the more important the mouth is for the CNN, the better is the performance of participants through the different experimental conditions. This result suggests that the statistical importance of different face features computed from CNN-based saliency maps could explain participants' performance within the different experimental conditions. Thus, it is likely that the participants use a similar statistical signal, i.e., a signal conveyed within the mouth area, to perform the task. The mouth importance systematically predicted which face on a pair was easier to detect (the emotional face), and predicted to a lesser extent which conditions facilitates this discrimination (e.g., in HSF than LSF, with a happy face).

Since some studies have shown that the eyes are important to categorise facial expressions, it may seem surprising that the eyes are not used by the CNN in this specific experiment. Indeed, several authors found that the eyes are important to recognise fear, whereas the mouth is important to recognise happiness (e.g.,Dailey et al., 2002; Eisenbarth and Alpers, 2011; M. L. Smith et al., 2005), and bottom-up saliency maps showed that the eye region was highly salient. In fact, our results do not exclude that the eyes are important for some tasks, for example, for the categorisation of fearful faces. But in the specific task of discriminating emotional from neutral faces, with a happy or a fearful emotional face, the mouth may be more informative. This result is supported by the fact that in our previous study we found that the eye saliency was unrelated to performance (Entzmann et al., 2022). Also, other studies suggest that the mouth alone might be sufficient to detect happy or angry expressions (e.g., Horstmann et al., 2012), or is used to a greater extent than the eyes when categorising facial expressions (Saumure

et al., 2018). Overall, it is not so surprising that the eyes play a limited role in our experiments, as the usefulness of the different features is highly dependent on the task to perform and the expressions that are used.

Also, as the goal of our model was to simulate the behavioural experiment, we limited our neural network to the use of happy and fearful emotional expressions. The use of different facial expressions could lead to different results, especially for CNN saliency maps. Indeed, happy faces are known for their diagnostic mouth, whereas the reliance on the eye and the mouth is more balanced for fearful faces. Using for example angry and fearful expressions, two expressions related to the eye region, may lead to a higher reliance on the eye region. Also, even if we have shown that statistical differences within the mouth can explain human performance, we do not exclude the possibility that a more complex processing is involved in humans. Then, further work will also be relevant to identify the other factors that may explain how humans discriminate an emotional from a neutral face.

With a very simple architecture, the CNN achieved high performance in discriminating emotional and neutral faces. Looking only at the global performance of the network (i.e., the accuracy of the categorisation), in the different spatial frequencies, with a happy or a fearful face, the model was difficult to compare with human performance in Entzmann et al. (2022). In the current study, the CNN outperforms human saccadic responses in each spatial frequency and emotion. We assume that this lack of similarity is explained by the fact that the model did not simulate the constraint of a saccadic response. For example, in the behavioural experiment, participants could not freely explore the images. They had to respond while fixating the centre of the screen, whereas the model had a precise access to every pixel on the images. Also, whereas saccadic responses are well suited for studying the time course of visual processing and attentional effects, they are known to lead to lower accuracy than manual responses (e.g., Bannerman et al., 2009; Kirchner and Thorpe, 2006). This can be due to the fact that participants have less control on eye movements, and can initiate saccades before a conscious decision is made. The use of manual instead of saccadic responses in the behavioural experiment could have made the network results more comparable to those of humans.

Simple linear regression analysis revealed that the RMS_{CNN} offered significant predictions of human accuracy, that were slightly better than predictions obtained with the $\text{RMS}_{Bottom-up}$ (with an adjusted R^2 of 0.76 and 0.74 for the RMS_{CNN} and $\text{RMS}_{Bottom-up}$ respectively). Then, incorporating task-related saliency led to an improvement of the predictions. However, we could have expected this difference to be larger. Looking only at edge and contrast differences around the mouth is sufficient to generate good predictions, without considering task-related saliency. Still, this result is not limiting the interest of CNN-based saliency maps. Indeed, the main advantage of the CNN-based saliency maps is that they directly highlight the useful features. In contrast, predictions from bottom-up saliency maps requires a knowledge about which region is informative for the task to generate predictions. Although we hypothesised that saccade endpoints can be predicted from both the RMS_{CNN} and $\text{RMS}_{Bottom-up}$, such evidence was only found for the RMS_{CNN}. Also, whereas the $\text{RMS}_{CNN}$was a significant predictor and was better than the $\text{RMS}_{Bottom-up}$, the quality of the predictions was not that high (with an adjusted R^2 of 0.28). It means that 28% of the variance of saccade endpoints is explained by the model. Then, although this result is encouraging, we suggest that endpoint positions may be the results of more complex interference. For example, they could be more dependent on the salience of the different regions of the face and not only on that of the mouth.

Overall, we suggest that CNN-based saliency maps can be an interesting tool to study task-related saliency in behavioural experiments. Here we applied their used in the context of a saccadic choice task, but they may be suitable for a variety of different tasks. For example, in visual exploration task when participants should find a specific target. The current task was designed for rapid detection of emotions, and we can hypothesise that, when participants have more time to explore visual stimuli, CNN-based saliency maps can predict human fixations better than bottom-up saliency maps (as highlighted byMurabito et al., 2018). Further studies in different contexts could lead to a better understanding of the efficiency of CNN-based saliency map to predict attention in humans.

Data availability statement

The data that support the findings of this study, and our codes, are available in the Open Science Framework repository at https://osf.io/vq3jy/.

Acknowledgments

This work was supported by NeuroCoG IDEX UGA in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02). This work has been partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003) to Martial Mermillod.

References

Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2018). Sanity checks for saliency maps. *Advances in neural information processing systems*, *31*.

Adolphs, R. (2008). Fear, faces, and the human amygdala. *Current opinion in neurobiology*, *18*(2), 166–172.

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual cognition*, *17*(6-7), 1185–1204.

Bannerman, R. L., Milders, M., & Sahraie, A. (2009). Processing emotional stimuli: Comparison of saccadic and manual choice-reaction times. *Cognition & Emotion*, *23*(5), 930–954. https://doi.org/10.1080/02699930802243303

Borji, A., & Itti, L. (2012). State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 185–207.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, *30*(6), 1081–1106. https://doi.org/10.1080/02699931.2015.1049124

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience*, *14*(8), 1158–1173. https://doi.org/10.1162/089892902760807177

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion*, *11*(4), 860–865. https://doi.org/10.1037/a0022758

Entzmann, L., Guyader, N., Kauffmann, L., Lenouvel, J., Charles, C., Peyrin, C., Vuillaume, R., & Mermillod, M. (2021). The role of emotional content and perceptual saliency during the programming of saccades toward faces. *Cognitive Science*, *45*(10), e13042.

Entzmann, L., Guyader, N., Kauffmann, L., Peyrin, C., & Mermillod, M. (2022). Detection of emotional faces: The role of spatial frequencies and local features. https://doi.org/10.31234/osf.io/b6n98

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. https://doi.org/10.1016/S0042-6989(01)00097-9

Horstmann, G., Lipp, O. V., & Becker, S. I. (2012). Of toothy grins and angry snarls–open mouth displays contribute to efficiency gains in search for emotional faces. *Journal of Vision*, *12*(5), 7–7. https://doi.org/10.1167/12.5.7

Huang, Y., Chen, F., Lv, S., & Wang, X. (2019). Facial expression recognition: A survey. *Symmetry*, *11*(10), 1189. https://doi.org/10.3390/sym11101189

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Jiao, Y., Niu, Y., Zhang, Y., Li, F., Zou, C., & Shi, G. (2019). Facial attention based convolutional neural network for 2d+3d facial expression recognition. *2019 IEEE Visual Communications and Image Processing (VCIP)*, 1–4. https://doi.org/10.1109/VCIP47243.2019.8965843

Jospin, L. V., Buntine, W., Boussaid, F., Laga, H., & Bennamoun, M. (2020). Hands-on bayesian neural networks–a tutorial for deep learning users. *arXiv preprint arXiv:2007.06823*.

Kartali, A., Roglic, M., Barjaktarovic, M., Duric-Jovicic, M., & Jankovic, M. M. (2018). Real-time algorithms for facial emotion recognition: A comparison of different approaches. *2018 14th Symposium on Neural Networks and Applications (NEUREL)*, 1–4. https://doi.org/10.1109/NEUREL.2018.8587011

Khanzada, A., Bai, C., & Celepcikay, F. T. (2020). Facial expression recognition with deep learning, 6.

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv:1412.6980 [cs]*. Retrieved November 18, 2019, from http://arxiv.org/abs/1412.6980

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776. https://doi.org/10.1016/j.visres.2005.10.002

Kruthiventi, S. S. S., Gudisa, V., Dholakiya, J. H., & Babu, R. V. (2016). Saliency unified: A deep architecture for simultaneous eye fixation prediction and salient object segmentation. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5781–5790. https://doi.org/10.1109/CVPR.2016.623

Kümmerer, M., Theis, L., & Bethge, M. (2014). Deep gaze i: Boosting saliency prediction with feature maps trained on imagenet. *arXiv preprint arXiv:1411.1045*.

Kümmerer, M., Wallis, T. S., & Bethge, M. (2016). Deepgaze ii: Reading fixations from deep features trained on object recognition. *arXiv preprint arXiv:1610.01563*.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Li, R., & Cottrell, G. (2012). A new angle on the EMPATH model: Spatial frequency orientation in recognition of facial expressions, 7.

Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 1–1. https://doi.org/10.1109/TAFFC.2020.2981446

Lundqvist, D., Flykt, A., & Ohman, A. (1998). The karolinska directed emotional faces (KDEF).

Mahdi, A., Qin, J., & Crosby, G. (2019). Deepfeat: A bottom-up and top-down saliency model based on deep features of convolutional neural networks. *IEEE Transactions on Cognitive and Developmental Systems*, *12*(1), 54–63.

Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y. H., Vuilleumier, P., Gil-Nagel, A., & Strange, B. A. (2016). A fast

pathway for fear in human amygdala. *Nature Neuroscience*, *19*(8), 1041–1049. https://doi.org/10.1038/nn.4324

Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation. *Neurocomputing*, *73*(13), 2522–2531. https://doi.org/10.1016/j.neucom.2010.06.002

Mermillod, M., Vermeulen, N., Lundqvist, D., & Niedenthal, P. M. (2009). Neural computation as a tool to differentiate perceptual from emotional processes: The case of anger superiority effect. *Cognition*, *110*(3), 346–357. https://doi.org/10.1016/j.cognition.2008.11.009

Minaee, S., Minaei, M., & Abdolrashidi, A. (2021). Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors*, *21*(9), 3046.

Mollahosseini, A., Chan, D., & Mahoor, M. H. (2016). Going deeper in facial expression recognition using deep neural networks. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1–10. https://doi.org/10.1109/WACV.2016.7477450

Mulckhuyse, M. (2018). The influence of emotional stimuli on the oculomotor system: A review of the literature. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(3), 411–425. https://doi.org/10.3758/s13415-018-0590-8

Murabito, F., Spampinato, C., Palazzo, S., Giordano, D., Pogorelov, K., & Riegler, M. (2018). Top-down saliency detection driven by visual classification. *Computer Vision and Image Understanding*, *172*, 67–76.

Noudoost, B., Chang, M. H., Steinmetz, N. A., & Moore, T. (2010). Top-down control of visual attention. *Current Opinion in Neurobiology*, *20*(2), 183–190. https://doi.org/10.1016/j.conb.2010.02.003

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology*, *30*(10), 953–958. https://doi.org/10.1016/j.psyneuen.2005.03.019

Pan, J., Sayrol, E., Giro-I-Nieto, X., McGuinness, K., & OConnor, N. E. (2016). Shallow and deep convolutional networks for saliency prediction. *2016 IEEE Con-*

ference on Computer Vision and Pattern Recognition (CVPR), 598–606. https://doi.org/10.1109/CVPR.2016.71

Pitaloka, D. A., Wulandari, A., Basaruddin, T., & Liliana, D. Y. (2017). Enhancing CNN with preprocessing stage in automatic emotion recognition. *Procedia Computer Science*, *116*, 523–529. https://doi.org/10.1016/j.procs.2017.10.038

Rouast, P. V., Adam, M., & Chiong, R. (2019). Deep learning for human affect recognition: Insights and new developments. *IEEE Transactions on Affective Computing*, 1–1. https://doi.org/10.1109/TAFFC.2018.2890471

Saumure, C., Plouffe-Demers, M.-P., Estéphan, A., Fiset, D., & Blais, C. (2018). The use of visual information in the recognition of posed and spontaneous facial expressions. *Journal of vision*, *18*(9), 21–21.

Schurgin, M., Nelson, J., Iida, S., Ohira, H., Chiao, J., & Franconeri, S. (2014). Eye movements during emotion recognition in faces. *Journal of vision*, *14*(13), 14–14.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Deep inside convolutional networks: Visualising image classification models and saliency maps. *In Workshop at International Conference on Learning Representations*.

Smith, F. W., & Schyns, P. G. (2009). Smile through your fear and sadness: Transmitting and identifying facial expression signals over a range of viewing distances. *Psychological Science*, *20*(10), 1202–1208. https://doi.org/10.1111/j.1467-9280.2009.02427.x

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*(3), 184–189. https://doi.org/10.1111/j.0956-7976.2005.00801.x

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, *11*(10), 697–709. https://doi.org/10.1038/nrn2889

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Vig, E., Dorr, M., & Cox, D. (2014). Large-scale optimization of hierarchical features for saliency prediction in natural images. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2798–2805. https://doi.org/10.1109/CVPR.2014.358

Vyas, A. S., Prajapati, H. B., & Dabhi, V. K. (2019). Survey on face expression recognition using CNN. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 102–106. https://doi.org/10.1109/ICACCS.2019.8728330

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*(9), 1395–1407. https://doi.org/10.1016/j.neunet.2006.10.001

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. how individual face parts contribute to successful emotion recognition. *PloS one*, *12*(5), e0177239.

Zelinsky, G., Zhang, W., Yu, B., Chen, X., & Samaras, D. (2005). The role of top-down and bottom-up processes in guiding eye movements during visual search. *Advances in neural information processing systems*, *18*.

Zheng, Q., Jiao, J., Cao, Y., & Lau, R. W. (2018). Task-driven webpage saliency. *Proceedings of the European conference on computer vision (ECCV)*, 287–302.

Chapitre 3 - Points clés

- Mise en place d'une expérience comportementale, d'une analyse de la saillance *bottom-up* et d'une simulation (CNN).
- La détection et le déclenchement de saccades vers des visages émotionnels ou neutres sont plus rapides et induisent moins d'erreurs en HFS qu'en BFS.
- La détection plus rapide des visages émotionnels et neutres en HFS qu'en BFS est seulement significative avec un contraste égalisé (interaction marginale entre les fréquences spatiales et le contraste).
- Les cartes de saillance issues d'un CNN peuvent être utilisées pour visualiser les régions utiles à une tâche.
- La région de la bouche est cruciale dans la détection de visages émotionnels ou neutres.
- La saillance de la bouche, qu'elle soit calculée à partir de cartes de saillance *bottom-up* ou des cartes de saillance issues du CNN, prédit significativement les performances des participants dans les différentes conditions expérimentales.

Chapitre 4

Implication de la voie sous-corticale dans la perception des visages apeurés : étude de l'activité en IRMf

Table des matières

4.1 Préface

Dans ce chapitre, nous nous sommes intéressés aux bases cérébrales de la discrimination des expressions faciales, ainsi qu'à leur lien avec le traitement des fréquences spatiales. Tout au long de ce travail de thèse, nous avons fondé nos hypothèses comportementales sur l'existence d'une voie sous-corticale, qui permettrait une détection rapide des visages émotionnels, en particulier apeurés, à partir des BFS. Cette hypothèse est notamment corroborée par une étude de Vuilleumier et al. (2003) en IRMf qui a montré des activations plus fortes en réponse à des visages apeurés (en comparaison à des visages neutres) dans l'amygdale et le thalamus (plus précisément au niveau du pulvinar ; un cluster d'activation qui s'étendait dans leur étude vers le CS). Cet effet était seulement observé lorsque les images étaient présentées en BFS ou non filtrées (voir la Figure 1.17 du premier chapitre). Dans le chapitre précédent, nous n'avons pas observé de capture de l'attention plus importante pour des visages apeurés, en BFS, en comparaison à d'autres visages (par exemple joyeux, en HFS). Bien que ce résultat remette en question l'aspect automatique de la capture de l'attention par ces visages, il ne remet pas en question le fait qu'ils puissent activer la voie sous-corticale plus fortement que les visages neutres. Ici, nous présentons

une étude effectuée en IRMf, dont le but était de reproduire les résultats de l'étude de Vuilleumier et al. (2003), mais également d'évaluer l'impact de l'égalisation du contraste sur cet effet. Ainsi, dans cette expérience, nous n'avons pas utilisé le paradigme de choix saccadique, mais un paradigme de catégorisation de genre, avec une réponse manuelle.

Il est important avant d'introduire cette étude de replacer son contexte, qui diffère de celui des études présentées dans les précédents chapitres. En effet, cette expérience a été effectuée dans le cadre d'un projet de plus grande envergure, le projet *EyeProxy*, qui réunit plusieurs chercheurs et laboratoires grenoblois impliqués dans l'étude des systèmes visuel et oculomoteur. Ce projet a pour objectif d'identifier les bases cérébrales de l'analyse visuelle et de la programmation de mouvements oculaires en enregistrant conjointement des signaux oculométriques et neuraux (IRMf et EEG). Il est prévu que les données obtenues sur des sujets sains soient partagées, afin de constituer une base de données qui pourra ensuite être comparée à des données obtenues chez des patients. Les participants recrutés dans ce projet ont pris part à 3 sessions expérimentales : 1 session en IRMf et oculométrie et 2 sessions en EEG et oculométrie. L'expérience présentée ici fait partie de la première session expérimentale, effectuée en IRMf. Lors de cette session, les participants ont réalisé plusieurs expériences réparties sur 1 heure : 3 expériences d'intérêt, incluant celle qui est présentée dans ce chapitre, et 2 expériences qui avaient pour but de localiser des régions spécifiques, notamment les aires sélectives aux visages. L'expérience présentée ici a aussi fait l'objet d'un enregistrement en EEG lors de la session 2 à laquelle participaient tous les sujets inclus dans *EyeProxy*. Néanmoins, seules les données IRMf sont présentées dans ce manuscrit.

4.2 Introduction

Le traitement des visages fait intervenir un réseau neural vaste, qui doit permettre de détecter les informations pertinentes, notamment les expressions faciales, rapidement pour que l'organisme puisse fournir une réponse adaptée. Bien que les visages attirent rapidement l'attention, les visages avec une expression émotionnelle sont détectés et attirent le regard plus rapidement que les visages neutres (Bannerman, Milders, de Gelder et al., 2009 ; Bannerman, Milders et Sahraie, 2009 ; Calvo et Nummenmaa, 2008 ; Entzmann et al., 2021 ; Frischen et al., 2008). Les corrélats cérébraux associés à cet effet peuvent être mis en évidence par des enregistrements de l'activité en IRMf, en identifiant les régions qui s'activent plus fortement face à des visages émotionnels que neutres. Comme nous l'avons vu dans le Chapitre 1, différentes expressions émotionnelles, en comparaison à des expressions neutres, activent des réseaux distincts, mais pas totalement séparables (Fusar-Poli et al., 2009 ; Sabatinelli et al., 2011 ; Vytal et Hamann, 2010). De manière générale, les expressions émotionnelles induisent, en comparaison à des expressions neutres, une augmentation de l'activité au niveau de l'amygdale, du gyrus fusiforme ou du gyrus occipital (Liu et al., 2021).

Certains modèles du traitement des stimuli émotionnels suggèrent que la détection de la pertinence émotionnelle soit basée sur un traitement rapide des BFS. Plus précisément,

le traitement des émotions, incluant le traitement des expressions faciales, se ferait par l'intermédiaire de deux voies distinctes : une voie sous-corticale pour la vision non consciente et une voie corticale pour la vision consciente (voir par exemple Tamietto et de Gelder, 2010). La voie sous-corticale transmettrait l'information de la rétine à l'amygdale, en passant par le pulvinar et le CS et en contournant ainsi le cortex visuel primaire. La voie corticale transmettrait l'information à l'amygdale plus tardivement, en passant par le cortex visuel primaire. Ainsi, la voie sous-corticale traiterait l'information grossière, en BFS, très rapidement, ce qui permettrait à l'organisme de réagir en présence d'un danger. L'information détaillée, en HFS, serait quant à elle traitée par la voie corticale. Une activation précoce de l'amygdale pourrait aussi guider le traitement de l'information émotionnelle dans la voie corticale (Gschwind et al., 2012 ; Sabatinelli et al., 2009).

Cette hypothèse d'un traitement en double voie qui différencierait le traitement des HFS et des BFS est en particulier corroborée par une étude de Vuilleumier et al. (2003), en IRMf, qui a montré des activations plus fortes en réponse à des visages apeurés (en comparaison à des visages neutres) au niveau de l'amygdale et du thalamus (plus précisément au niveau du pulvinar ; un cluster d'activation qui s'étendait dans leur étude vers le CS). Cet effet était seulement observé lorsque les images étaient présentées en BFS ou non filtrées. Ils ont aussi mis en avant une activité plus forte pour les HFS que les BFS (et pour les visages apeurés que neutres, mais seulement en BFS) au niveau du gyrus fusiforme. Des résultats similaires ont été observés avec un enregistrement en EEG intracrânien. Plus précisément, Méndez-Bértolo et al. (2016) ont mis en évidence des activations plus fortes pour des visages apeurés que neutres dès 74 ms après l'apparition du stimulus. Cet effet n'était présent que lorsque les images étaient présentées en BFS ou non filtrées, et émergeait de la partie latérale de l'amygdale. Plusieurs autres études se sont intéressées au traitement cérébral des émotions en fonction des fréquences spatiales. En EEG, en accord avec les résultats cités précédemment, plusieurs études ont mis en avant une différence précoce entre les expressions faciales, mais pas lorsque les visages étaient présentés en HFS (Nakashima et al., 2008 ; Pourtois et al., 2005 ; Vlamings et al., 2009). Aussi, Van Le et al. (2013) ont effectué des enregistrements en EEG intracrânien de l'activité du pulvinar en réponse à des images de serpents. Ils ont montré une activité plus forte lorsque les images étaient présentées en BFS plutôt qu'en HFS.

Néanmoins, plusieurs autres études n'observent pas une telle sélectivité de la voie sous-corticale aux BFS et aux émotions. Par exemple, Corradi-Dell'Acqua et al. (2014) ont analysé l'activité en IRMf en réponse à des images hybrides contenant les HFS d'un visage émotionnel et les BFS d'un visage neutre ou inversement, et cela chez des sujets sains ou des patients. Ils n'ont pas observé de différence entre les visages émotionnels et neutres en BFS chez les sujets sains. Dans une autre étude, l'activité de l'amygdale en réponse à des visages neutres ou apeurés, présentés en BFS, a été analysée, aussi chez des sujets sains ou des patients (Ottaviani et al., 2012). Les résultats n'ont pas révélé d'effet significatif de l'expression du visage sur l'activité de l'amygdale. Aussi, dans une étude pilote, Campagne et al. (2016) ont étudié les activations cérébrales en IRMf en réponse à des scènes visuelles sous différents filtrages. Ils ont observé des activations plus fortes

dans des régions comme l'amygdale et le gyrus fusiforme pour des stimuli émotionnels que neutres. En accord avec l'étude de Vuilleumier et al. (2003), leurs résultats ont montré que les HFS provoquaient une activité plus forte que les BFS au niveau du gyrus fusiforme. Néanmoins, une analyse en région d'intérêt sur l'amygdale a seulement révélé un effet de la valence émotionnelle, sans interaction avec les fréquences spatiales. Pour finir, certaines études ont utilisé la modélisation causale dynamique pour évaluer la connectivité entre les régions de la voie sous-corticale et corticale lors de la perception de visages exprimant différentes émotions (Garvert et al., 2014 ; McFadyen et al., 2017). Dans ces études, les auteurs avaient notamment comparé la probabilité de modèles comportant seulement une voie corticale (reliant le LGN, V1 et l'amygdale), à celle de modèles comportant à la fois une voie corticale et une voie sous-corticale (reliant le pulvinar et l'amygdale), en se basant sur des enregistrements de l'activité en MEG. D'après leurs résultats, les modèles incluant la voie sous-corticale sont les plus probables. Néanmoins, cette voie ne serait influencée ni par les émotions du visage ni par les fréquences spatiales.

Ainsi, les résultats de ces différentes études sont assez hétérogènes. Dans l'étude de McFadyen et al. (2017), les auteurs suggèrent que l'absence de sensibilité aux fréquences spatiales et aux émotions de la voie sous-corticale pourrait s'expliquer par le fait qu'ils utilisent des images égalisées en contraste. En fait, naturellement, après un filtrage des fréquences spatiales, les BFS et les HFS diffèrent, non seulement en termes de contenu fréquentiel, mais aussi en termes de contraste. Cela vient du fait que, dans le spectre d'amplitude des images, l'énergie décroît avec l'augmentation des fréquences spatiales (Field, 1987 ; Loftus et Harley, 2005 ; Tolhurst et al., 1992 ; van der Schaaf et van Hateren, 1996). Ainsi, dans les études où le contraste n'est pas égalisé après le filtrage, le contraste est plus important dans les BFS que dans les HFS. Lors de la comparaison de données comportementales ou neurophysiologiques, le contraste pourrait à lui seul expliquer les différences observées entre les BFS et les HFS.

Le contraste plus élevé des BFS pourrait favoriser leur traitement, car les objets plus contrastés sont généralement mieux détectés (Lupp et al., 1976 ; Vassilev et Mitov, 1976) ou catégorisés (Perfetto et al., 2020 ; Vlamings et al., 2009). Dans le cadre de la catégorisation d'expressions faciales (neutres ou apeurées), Vlamings et al. (2009) ont observé une catégorisation plus rapide d'images présentées en BFS qu'en HFS, mais cet effet était réduit lorsque le contraste était égalisé. Des résultats similaires ont été obtenus avec des images de scènes (Perfetto et al., 2020). Au niveau neural, le contraste pourrait aussi avoir un effet positif sur l'activation de plusieurs zones impliquées dans la perception visuelle, notamment V1 et l'amygdale (Boynton, 2005 ; Goodyear et Menon, 1998 ; Inagaki et al., 2012). Dans le contexte de la perception de scènes, Kauffmann, Ramanoël et al. (2015) ont observé des activations plus fortes dans l'aire parahippocampique des lieux pour des images en BFS (ou non filtrées) qu'en HFS ; cet effet était inversé lorsque les images étaient égalisées en contraste. Dans l'étude de McFadyen et al. (2017), le contraste était égalisé, ce qui n'était pas le cas dans d'autres études qui ont rapporté un effet des fréquences spatiales et des émotions (Méndez-Bértolo et al., 2016 ; Vuilleumier et al., 2003).

Dans l'expérience présentée dans ce chapitre, l'objectif était d'évaluer les bases neurales de la discrimination entre un visage neutre et apeuré, en fonction des fréquences spatiales, en mesurant également l'effet de l'égalisation du contraste des images filtrées. De manière similaire à ce qui a été fait par Vuilleumier et al. (2003), nous avons présenté à des participants des visages neutres ou apeurés filtrés tout en procédant à un enregistrement de l'activité en IRMf (et en oculométrie). Les images pouvaient être présentées en HFS ou en BFS, et soit égalisées en contraste, soit non égalisées en contraste. Pour répondre aux contraintes temporelles de l'expérience, nous n'avons pas inclus d'images non filtrées. Les visages étaient présentés au centre de l'écran et les participants devaient catégoriser le genre des visages. Dans la condition de contraste non égalisé, nous nous attendions à reproduire les résultats de l'étude de Vuilleumier et al. (2003). Plus précisément, nous nous attendions à une activité plus forte pour les HFS que les BFS au niveau du gyrus fusiforme (dans la zone qui correspond à la FFA) et à une activité plus forte pour les visages apeurés que neutres, mais seulement pour les BFS. Ensuite, nous nous attentions à une activité plus forte pour les visages apeurés que neutres au niveau de l'amygdale, du pulvinar et du CS, mais seulement pour les images présentées en BFS. Nous nous attendions à observer cette interaction entre les émotions et les fréquences spatiales seulement dans la condition de contraste non égalisé. En effet, nous supposions que l'égalisation du contraste pourrait inhiber ces différences, et ainsi expliquer l'absence d'effet observé dans l'étude de McFadyen et al. (2017).

4.3 Méthode

4.3.1 Participants

Au total, quarante-cinq participants volontaires ont été inclus dans le projet *EyeProxy*. Dans le cadre de cette expérience, les données de sept participants n'ont pas pu être acquises pour cause de problèmes techniques lors de l'enregistrement. Par ailleurs, les données de treize participants n'ont pas été incluses dans l'analyse, car le placement de la fenêtre d'acquisition n'englobait pas la totalité de l'amygdale [1]. Ainsi, vingt-cinq participants ont finalement été inclus dans cette analyse (13 femmes ; $M \pm SD$: 26 ± 1.2 ans ; tranche d'âge : 21-38 ans). Tous les participants avaient une vision normale ou corrigée à la normale (par l'intermédiaire du port de lunettes ou de lentilles), et aucun d'entre eux n'a reporté de maladie psychiatrique ou neurologique passée ou présente. À l'issue des trois sessions expérimentales du projet *EyeProxy*, ils recevaient une indemnisation à hauteur de 100 euros. Tous les participants ont également effectué une visite médicale autorisant leur inclusion dans le projet, et ont donné leur

1. Dans le cadre du projet *EyeProxy*, afin d'optimiser la résolution spatiale des images, la fenêtre d'acquisition ne couvrait pas la totalité du cerveau. Elle couvrait une zone allant du lobe frontal au lobe occipital. L'amygdale était située dans la partie inférieure de la fenêtre. Néanmoins, pour certains participants, le manipulateur radio ne pouvait pas couvrir toutes les zones d'intérêt correctement et la fenêtre d'acquisition coupait une partie de l'amygdale. Les données de ces participants n'ont pas été incluses dans notre analyse. Une visualisation des zones couvertes par l'enregistrement est présentée dans la Figure 4.3.

consentement éclairé, en conformité avec le code de conduite éthique de l'Association Médicale Mondiale (Déclaration de Helsinki) pour les expériences impliquant la personne humaine. Cette expérimentation clinique est régie par la loi française (Jardé, Décret n°2016-1537 16/11/2016 du 17/11/2016). L'ensemble des expériences menées dans le cadre du projet *EyeProxy* a été approuvé par un comité d'éthique (Comité de Protection des Personnes Sud-Est III, Eudra-CT 2020-100503-36).

4.3.2 Stimuli

Les stimuli utilisés dans cette expérience ont été construits à partir de 120 photographies de visages de la base KDEF (Lundqvist et al., 1998). Ces photographies correspondaient à 60 acteurs différents, exprimant une expression neutre ou apeurée. La construction des stimuli a été réalisée en plusieurs étapes. D'abord, afin d'obtenir des images similaires à celles de l'étude de Vuilleumier et al. (2003), toutes les images ont été rognées. Plus précisément, tous les visages ont été enfermés dans un cadre rectangulaire de 350 x 251 pixels excluant la plupart des cheveux. Ensuite, la luminance moyenne ainsi que le contraste de luminance *root mean square* (RMS ; qui correspond à l'écart-type des valeurs de l'intensité des pixels) moyen de chaque image ont été égalisés, afin d'obtenir une luminance moyenne de 126 et un contraste RMS moyen de 47, pour des valeurs d'intensité des pixels comprises entre 0 et 255 (ces valeurs correspondaient aux valeurs moyennes obtenues en prenant en compte tous les stimuli). Puis, les images ont été filtrées, afin d'en obtenir une version en BFS et une version en HFS, avec des fréquences de coupure respectivement fixées à 6 cpi pour les stimuli en BFS et 24 cpi pour les stimuli en HFS. Ce filtrage a été réalisé deux fois, d'abord sans égalisation du contraste entre les fréquences, puis avec égalisation du contraste entre les fréquences. Après le filtrage, les images avaient toujours la même luminance moyenne (fixée à une valeur de 126). Dans la condition de contraste non égalisé, le contraste RMS moyen était de 63 pour les images en BFS, et 12 pour les images en HFS. Dans la condition de contraste égalisé, le contraste RMS moyen des images HSF et BSF était de 38, correspondant environ à la moyenne du contraste des HFS et des BFS dans la condition de contraste non égalisé. Dans la configuration de l'expérience, les images mesuraient 5 x 6.3 ° d'angle visuel. Au final, 480 images ont été utilisées, correspondant à une condition d'Expression Faciale Émotionnelle ou *Emotional Facial Expression* (EFE ; Apeurée ou Neutre), une condition de Fréquences Spatiales (BFS ou HFS) et une condition de Contraste (non égalisé ou égalisé ; NonEG ou EG). La Figure 4.1 présente des exemples de stimuli obtenus.

4.3.3 Procédure

L'expérience était divisée en trois scans fonctionnels de 5 minutes chacun, élaborés en utilisant un paradigme de type bloc. Matlab (MathWorks, Natick, MA) et la Psychophysics Toolbox (Brainard, 1997) ont été utilisés pour contrôler le timing et l'affichage des stimuli. Chaque scan était composé de 30 blocs, correspondants à 6 blocs de repos, de 10 secondes, et 24 blocs de tâche, de 9 ou 12 secondes. Au cours des blocs de tâche, les participants

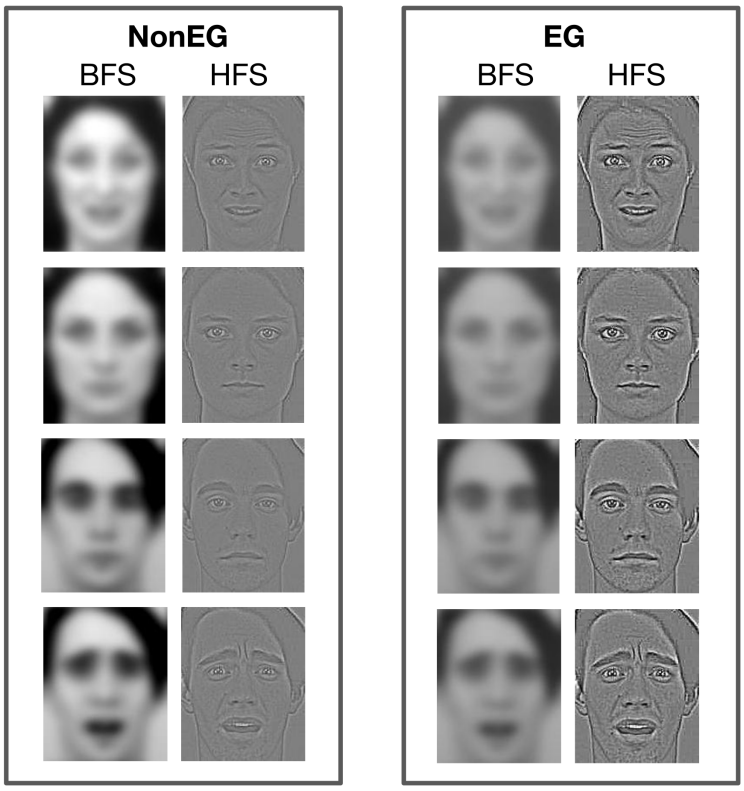

Figure 4.1 – Exemple de stimuli utilisés au cours de l'expérience, en condition de contraste non égalisé (NonEG, à gauche) ou égalisé (EG, à droite). Dans la condition NonEG, les images en BFS et HFS ont un contraste RMS moyen de 63 et 12, respectivement. Dans la condition EG, toutes les images ont un contraste RMS moyen de 38. La luminance moyenne a été fixée à 126 et était la même dans toutes les conditions (pour des valeurs d'intensité de pixel comprises entre 0 et 255).

devaient catégoriser le genre de visages sur plusieurs essais. Un essai se déroulait de la manière suivante. D'abord, une croix de fixation blanche était présentée sur un fond gris durant 500 ms. Ensuite, un stimulus était affiché pendant 200 ms, suivi d'un écran gris, affiché pendant 800 ms, qui terminait l'essai. Les participants avaient pour instruction d'appuyer sur un bouton du boîtier de réponse en fonction du genre du visage (masculin ou féminin). La correspondance des deux boutons avec le genre, masculin ou féminin, était contrebalancée entre les participants. L'exactitude de la réponse ainsi que le temps de réponse étaient enregistrés pour chaque essai. Chaque bloc de tâche correspondait à une condition expérimentale, c'est-à-dire à une condition de Fréquences Spatiales (HFS ou BFS), d'EFE (Apeurée ou Neutre) et de Contraste (NonEG ou EG). Dans chaque scan, il y avait trois blocs par condition : 2 blocs de 9 secondes (équivalent à 6 essais) et 1 bloc de 12 secondes (équivalent à 8 essais), et chaque bloc n'était jamais distant de plus de 100 secondes d'un bloc de la même condition. La Figure 4.2 présente une visualisation du déroulement d'un scan, dans laquelle chaque bloc d'une même condition est représenté par une même couleur. L'ordre des conditions était aléatoire. Avant l'expérience, les participants visualisaient quelques essais afin de se familiariser avec la tâche.

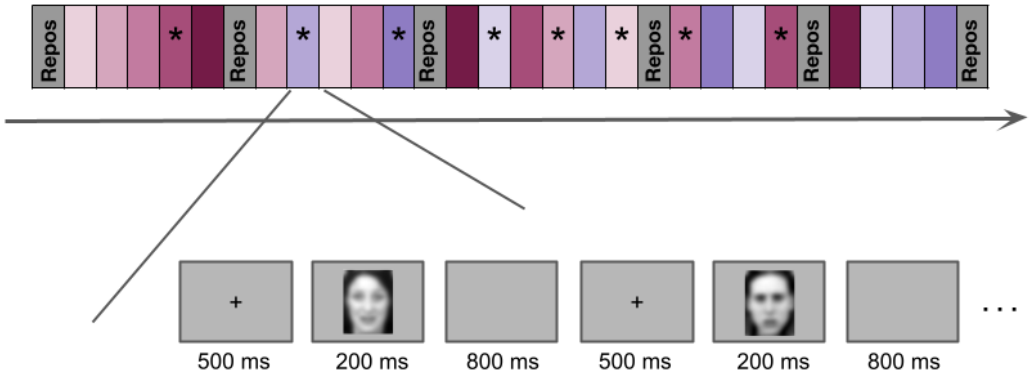

Figure 4.2 – Déroulement d'un scan fonctionnel. On a une alternance de blocs de repos (en gris) et de blocs de tâche (en nuances de violet). Chaque bloc de tâche correspond à une condition de Fréquences spatiales (HFS, BFS), d'EFE (Apeurée, Neutre) et de Contraste (EG, NonEG). Les blocs de tâche avec un marquage * durent 12 secondes, les autres 9 secondes. Les blocs de repos durent 12 secondes. Chaque bloc de tâche représente une succession d'essais. Un essai correspond à la présentation d'une croix de fixation durant 500 ms, puis d'un stimulus pendant 200 ms et finalement d'un écran gris pendant 800 ms. Les participants devaient appuyer sur une touche du boîtier de réponse à chaque essai en fonction du genre du visage (masculin ou féminin).

4.3.4 Données IRMf

4.3.4.1 Acquisition

Les données ont été acquises au centre hospitalier de Grenoble, sur la plateforme IRMaGe, à l'aide d'un scanneur IRM 3 Tesla (Achieva 3.0T TX Philips, Philips Medical Systems, Best, NL) et d'une antenne 32 canaux. Un miroir était placé sur l'antenne pour permettre aux participants de voir l'écran sur lequel étaient projetés les stimuli, situé à l'arrière de l'IRM. Lorsque les participants étaient installés dans l'aimant, un scan de repérage était d'abord effectué afin de régler la position de la fenêtre d'acquisition utilisée dans les scans suivants. Pour les scans fonctionnels, la fenêtre était orientée de manière à couvrir les structures impliquées dans les premières étapes du traitement visuel (notamment le CS, le CGL, V1 et le champ oculaire frontal). Afin d'optimiser la résolution spatiale des images acquises, elle excluait une partie du cerveau (voir la Figure 4.3 pour une visualisation du masque obtenu à partir du recouvrement des fenêtres d'acquisition de tous les participants). Après le scan de repérage, les participants réalisaient trois scans fonctionnels, indépendants de l'expérience présentée ici, dont le but était de localiser certaines régions d'intérêt. Ensuite, des volumes anatomiques du cerveau entier ont été acquis (un volume structurel et un volume avec les mêmes paramètres que les scans fonctionnels pour faciliter le réalignement des données fonctionnelles et structurelles). Puis, les participants réalisaient deux scans fonctionnels, encore indépendants de l'expérience présentée ici. Enfin, ils réalisaient les trois scans fonctionnels présentés dans le paragraphe précédent, qui font l'objet de ce chapitre. Ainsi, notre expérience débutait généralement 30 minutes après l'installation du participant dans l'IRM. Après ça, des images structurelles et des mesures de perfusions étaient également réalisées. Au total, l'acquisition des

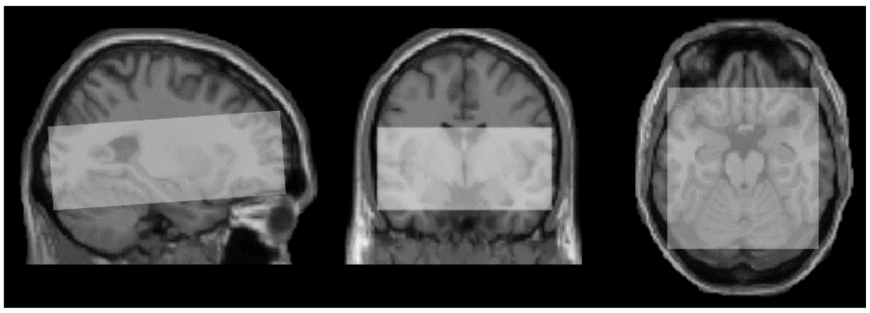

Figure 4.3 – Visualisation de la fenêtre d'acquisition correspondant à l'analyse des activations fonctionnelles, obtenue à partir du recouvrement des fenêtres d'acquisition de tous les participants inclus dans notre étude.

données IRM pour un participant durait 1h, dont 15 minutes étaient consacrées à notre expérience. Les images fonctionnelles (121 volumes par scan pour cette expérience) ont été acquises par l'intermédiaire de séquences pondérées en T2*, en écho de gradient rapide (*echo planar imaging*, EPI), avec les paramètres suivants : TR/TE = 2500/30 ms, angle de basculement = 80°, matrice d'acquisition = 80 x 157, champ de vue = 120 x 240 x 72, 48 coupes transversales, épaisseur des coupes = 1.35 mm, taille des voxels = 1.5 x 1.5 x 1.35 mm. Les images anatomiques ont été obtenues en utilisant une séquence pondérée en T1 (3D MP-RAGE) avec les paramètres suivants : résolution spatiale = 0.90 x 0.89 x 1 mm3, nombre de coupes sagittales = 220, matrice d'acquisition = 250 x 257 x 220, TR/TE/TI = 8.1/13.7/678 ms, angle de basculement = 8°. Pour stabiliser le champ magnétique, 4 mesures de volume (*dummies scans*) étaient réalisées avant chaque acquisition fonctionnelle.

4.3.4.2 Prétraitement

Avant la mise en place de l'analyse statistique, les données acquises ont été prétraitées à l'aide du logiciel *Statistical Parametric Mapping* 12 (SPM 12, Wellcome Department of Imaging Neuroscience, Londres, UK). Ce prétraitement peut se diviser en plusieurs étapes. D'abord, les volumes fonctionnels issus des différents scans ont été réalignés en fonction du volume le plus proche du volume anatomique, afin de corriger les déplacements causés par les mouvements de la tête (rotations et translations). À la fin de cette étape, une image correspondant à la moyenne des images fonctionnelles a été obtenue. Deuxièmement, les images fonctionnelles ont été mises en correspondance avec les images anatomiques. Plus précisément, le volume fonctionnel du cerveau entier a d'abord été recalé sur l'image moyenne obtenue à l'étape précédente (qui correspond à une fenêtre d'acquisition limitée, et non au cerveau entier). Puis, l'image anatomique a été recalée sur le volume fonctionnel du cerveau entier. Troisièmement, un lissage spatial gaussien a été appliqué sur les images fonctionnelles, pour corriger les différences interindividuelles et améliorer le rapport signal sur bruit. Ainsi, au cours de cette étape la valeur d'un voxel est remplacée par la moyenne pondérée par un noyau gaussien de 5 mm des valeurs de ses voxels voisins. Quatrièmement,

l'image anatomique a été segmentée de manière à classer les différents tissus cérébraux, et les champs de déformation qui permettent de passer de l'espace individuel à un espace normalisé ont été estimés. Pour finir, les images fonctionnelles et anatomiques ont été normalisées dans l'espace MNI (*Montreal Neurological Institute*), à l'aide des champs de déformations calculés précédemment, et elles ont été lissées de la même manière que les images dans l'espace du sujet. À l'issue du prétraitement, les images fonctionnelles et anatomiques étaient exprimées dans l'espace du sujet et dans l'espace MNI.

4.3.4.3 Analyses statistiques

Le traitement statistique a été réalisé à l'aide du logiciel SPM 12 (Wellcome Department of Imaging Neuroscience, Londres, UK). Des analyses de premier niveau ont été mises en place en utilisant le modèle linéaire général (Friston et al., 1994). Plus précisément, une concaténation des trois sessions a d'abord été réalisée à l'aide de la fonction spm-concatenate [2]. Puis, pour chaque participant, huit conditions d'intérêt ont été modélisées comme régresseurs. Nous avons dissocié 2 conditions de Fréquences Spatiales (BFS, HFS), 2 conditions d'EFE (Apeurée, Neutre) et 2 conditions de Contraste (NonEG, EG). Les six paramètres de mouvements obtenus durant le réalignement ont été pris en compte dans le modèle comme régresseurs de non-intérêt. Aussi, les données de chaque voxel ont été filtrées de manière à supprimer les dérives temporelles en basses fréquences (filtrage passe-haut avec une fréquence de coupure à 1/128 Hz). Ensuite, des analyses de second niveau ont été effectuées. Nous avons d'abord réalisé une analyse de type *full factorial* ($N = 25$, $k = 5$, $p < .001$, $F = 11.16$; Wellcome Department of Imaging Neuroscience, Londres, UK) afin d'étudier l'effet principal de l'EFE et des Fréquences Spatiales. Cette analyse avait pour but de générer des cartes des activations fonctionnelles (définies sur la base d'un seuil non corrigé $p <.001$, avec une taille de cluster $k \geq 5$ voxels) sur toute la zone d'acquisition, qui rendent compte des régions impliquées de manière générale dans le traitement des fréquences spatiales et des expressions faciales. Nous avons ensuite généré les cartes des activations correspondant à l'effet de l'EFE et des Fréquences Spatiales dans chaque condition expérimentale [3]. Les clusters obtenus ont été labellisés à l'aide de la toolbox *Automated Anatomical Labeling* 3 (AAL ; Tzourio-Mazoyer et al., 2002).

Nous avons ensuite effectué une analyse en région d'intérêt (ROI). Nous avons extrait le pourcentage de changement de signal (PCS) dans chacune de nos conditions expérimentales dans les régions qui nous intéressent particulièrement : la FFA, l'OFA, l'amygdale latérale et médiale, le pulvinar et le CS. Les masques correspondant à chaque ROI ont été définis dans l'espace MNI de manière différente pour chaque région. Pour l'OFA et la FFA, qui sont des régions sélectives aux visages connues pour discriminer les

2. Avant de procéder à l'estimation, cette fonction ajoute un régresseur tenant compte de l'effet des sessions et corrige le filtre passe-haut et les calculs de non-sphéricité temporelle pour tenir compte des longueurs des sessions originales.

3. Par exemple, pour l'effet des fréquences spatiales, les cartes ont été générées d'abord indépendamment du contraste et de l'expression faciale, puis dans chaque condition de contraste et d'expression faciale : avec un visage apeuré, en condition de contraste EG et en condition de contraste NonEG, et avec un visage neutre, en condition de contraste EG et en condition de contraste NonEG

expressions faciales (voir par exemple Kadosh et al., 2011 ; Liu et al., 2021 ; Pitcher et al., 2011 ; Xu et al., 2021), les masques ont été définis individuellement chez chaque participant en utilisant un scan fonctionnel indépendant [4]. Plus précisément, ils ont été extraits à partir de clusters obtenus sur les cartes des activations correspondant au contraste [Visage - Objet], définies sur la base d'un seuil non corrigé $p < .001$, avec une taille de cluster $k \geq 10$ voxels. Sur les 25 participants inclus dans cette analyse, la FFA a été obtenue pour 21 d'entre eux, et l'OFA pour 20 d'entre eux. Parfois, les clusters d'activations étaient localisés dans l'hémisphère gauche, parfois dans l'hémisphère droit et parfois de manière bilatérale (voir le Tableau C.1 en appendice qui détaille les activations pour chaque participant). Dans ce dernier cas, la région qui présentait le pic d'activation le plus fort au niveau du voxel (en termes de valeur statistique) était définie comme ROI. Au final, la FFA a été définie à gauche pour 7 participants, et l'OFA pour 5 participants. L'amygdale a été définie à droite et à gauche sur la base de l'atlas Brainnetome (Fan et al., 2016 ; https://atlas.brainnetome.org/). Nous avons dissocié la partie médiale de la partie latérale de l'amygdale. Pour le pulvinar et le CS, nos masques correspondaient à ceux utilisés dans l'étude de McFadyen et al. (2019) sur la connectivité entre le pulvinar, l'amygdale et le CS. Plus précisément, le pulvinar a été défini à droite et à gauche en utilisant la délimitation fonctionnelle de Barron et al. (2015), dans laquelle plusieurs sous-régions du pulvinar sont distinguées. Le masque utilisé ici correspond à la fusion de ces différentes sous-régions. Pour le CS, le masque était défini anatomiquement dans l'espace MNI. Une visualisation des différentes ROI est présentée sur la Figure 4.4. L'analyse statistique du PCS moyen dans nos ROI a été mise en place à l'aide du logiciel R (R Core Team, 2016). Une ANOVA à mesures répétées avec les Fréquences Spatiales (BFS, HFS), l'EFE (Apeurée, Neutre) et le Contraste (EG, NonEG) comme facteurs intra-sujet a été appliquée sur les PCS moyens de chaque participant pour chaque ROI. Si nécessaire (c'est-à-dire si un effet d'interaction était observé), des tests t à échantillons appariés étaient utilisés pour les comparaisons par paires, et une correction de Bonferroni était appliquée pour corriger le seuil de significativité des comparaisons multiples. Les tailles d'effet ont été estimées en calculant l'êta-carré partiel (η_p^2), et un effet était considéré comme significatif si sa valeur p était inférieure au seuil $\alpha = .05$. Les effets marginalement significatifs (c'est-à-dire dont la valeur p se situe entre .05 et .1) ne sont pas reportés.

4.3.5 Données comportementales

Dans l'expérience, la réponse et le temps de réponse étaient enregistrés pour chaque essai. Ces variables ont également été analysées à l'aide du logiciel R (R Core Team, 2016). Plus précisément, nous avons analysé pour chaque participant la proportion de réponses correctes et le temps de réaction moyen pour les réponses correctes dans chaque condition expérimentale. Ainsi, une ANOVA à mesures répétées a été mise en place, avec l'EFE (Apeurée, Neutre), les Fréquences Spatiales (HFS, BFS) et le Contraste (EG,

4. Scan fonctionnel de 4 minutes réalisé avant notre expérience d'intérêt, pendant lequel les participants voyaient des images de visages et d'objets. Ce scan a été mis en place dans le but précis de localiser les régions sélectives aux visages. Voir l'Appendice C.1 pour plus de détails.

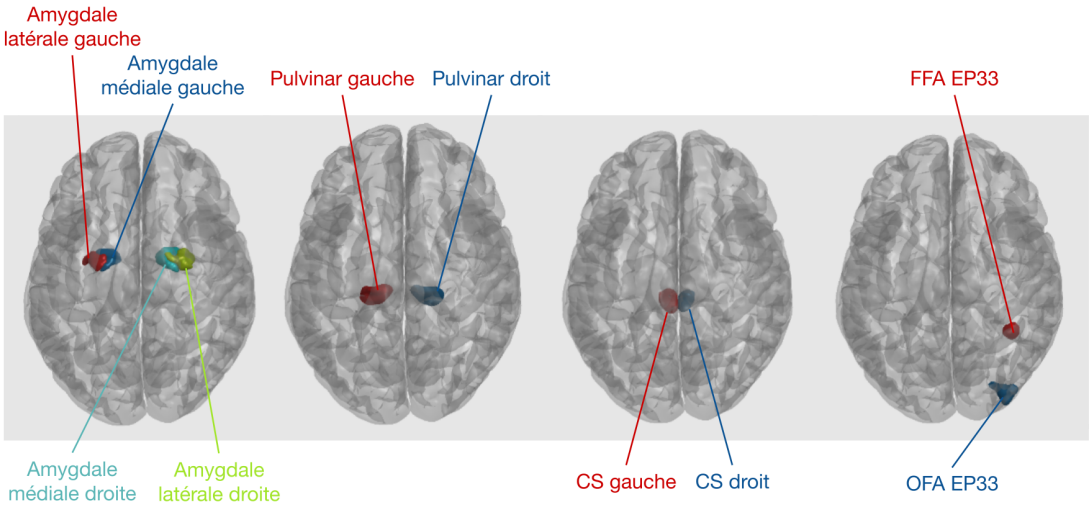

Figure 4.4 – Visualisation des masques correspondant aux différentes régions d'intérêt. Pour la FFA et l'OFA, les masques sont définis pour chaque participant individuellement. Dans cette visualisation, nous présentons à titre d'exemple les masques obtenus pour un participant (EP33).

NonEG) comme facteurs intra-sujet. Si nécessaire (c'est-à-dire si un effet d'interaction était observé), des tests t à échantillons appariés étaient utilisés pour les comparaisons par paires, et une correction de Bonferroni était appliquée. Les tailles d'effet ont été estimées en calculant l'êta-carré partiel (η_p^2) pour les ANOVA et le d de Cohen pour les tests t. Un effet était considéré comme significatif si sa valeur p était inférieure au seuil $\alpha = .05$. Les effets marginalement significatifs (c'est-à-dire dont la valeur p se situe entre .05 et .1) ne sont pas reportés.

4.4 Résultats

La présentation des résultats est organisée de la façon suivante. D'abord, les résultats comportementaux (proportion de réponses correctes et temps de réaction) sont présentés. Ensuite, les cartes des activations correspondant à l'effet de l'EFE et à l'effet des Fréquences Spatiales sont décrites. Nous détaillons dans un premier temps les clusters obtenus de manière globale, indépendamment des conditions expérimentales. Puis, dans un second temps nous détaillons les clusters obtenus dans chaque condition expérimentale. Pour finir, les résultats de l'analyse en ROI sont présentés.

4.4.1 Résultats comportementaux

4.4.1.1 Proportion de réponses correctes

L'ANOVA à mesures répétées effectuée sur les proportions de réponses correctes moyennes (Figure 4.5-a) a révélé un effet principal de l'EFE ($F(1,19) = 5.6$, $p = .029$, $\eta_p^2 = .006$) et des Fréquences Spatiales ($F(1,19) = 5.5$, $p = .03$, $\eta_p^2 = .002$). Les performances étaient plus élevées lorsque les visages étaient neutres ($M \pm SD$: $.86 \pm .047$) plutôt

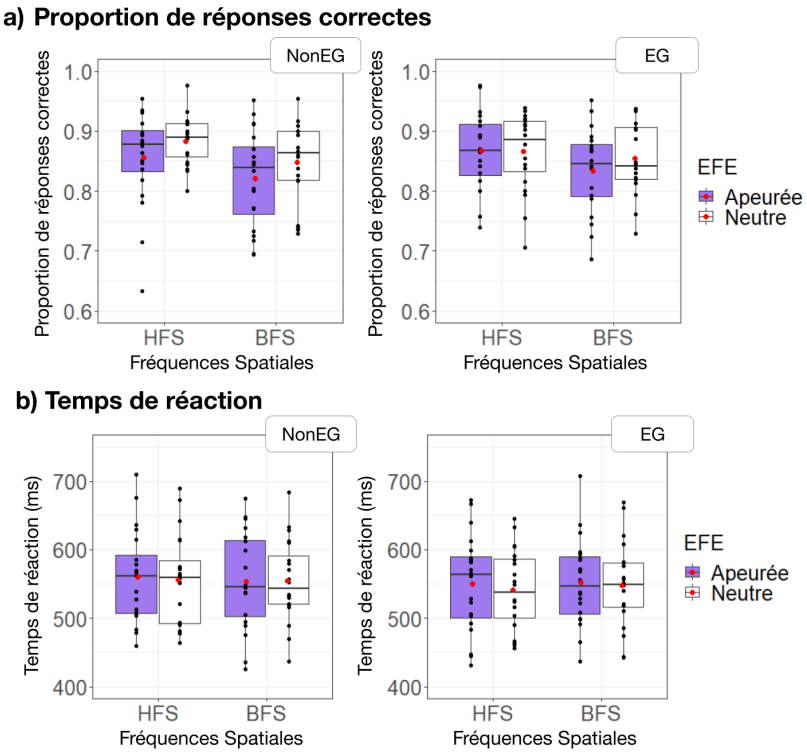

Figure 4.5 – Résultats comportementaux. (a) Proportion de réponses correctes et (b) latence des réponses correctes en condition de contraste non égalisé (à gauche) ou égalisé (à droite) en fonction de l'Expression Faciale Émotionnelle du visage (Apeurée, Neutre) et des Fréquences Spatiales (HFS, BFS).

qu'apeurés ($M \pm SD$: .84 ± .05), et en HFS ($M \pm SD$: .87 ± .048) plutôt qu'en BFS ($M \pm SD$: .84 ± .058).

4.4.1.2 Temps de réaction

L'ANOVA à mesures répétées effectuée sur les temps de réaction moyens (Figure 4.5-b) a révélé un effet principal du Contraste ($F(1,19) = 7.6$, $p = .013$, $\eta_p^2 = .001$). Les temps de réaction étaient plus faibles lorsque les images étaient égalisées en contraste ($M \pm SD$: 545 ± 66 ms) plutôt que non égalisées ($M \pm SD$: 553 ± 67 ms).

4.4.2 Cartes des activations fonctionnelles

4.4.2.1 Effet de l'expression faciale émotionnelle

Le contraste [Apeurée - Neutre], effectué indépendamment des conditions de Fréquences Spatiales (BFS, HFS) et de Contraste (EG, NonEG), a révélé des activations plus fortes dans des régions occipito-temporales, incluant les gyri fusiformes droit et gauche, le gyrus temporal moyen droit et gauche, le gyrus temporal supérieur droit et le gyrus occipital inférieur gauche. L'insula gauche, l'amygdale gauche et l'OFC postérieur droit étaient également plus activés pour les visages apeurés que neutres.

Le contraste [Apeurée - Neutre], effectué dans chaque condition de Fréquences Spatiales et de Contraste, a révélé les activations suivantes. En BFS, avec un contraste non égalisé (NonEG), le gyrus fusiforme gauche, le putamen gauche ainsi que le gyrus temporal supérieur étaient impliqués. En BFS, avec un contraste égalisé (EG), le réseau impliqué était un peu plus large. Il comprenait le gyrus fusiforme gauche, le gyrus temporal inférieur droit, le gyrus occipital supérieur gauche, l'OFC postérieur droit, le putamen droit, l'hippocampe droit et gauche ainsi que le noyau ventral latéral du thalamus gauche. En HFS, avec un contraste non égalisé (NonEG), le contraste [Apeurée - Neutre] a révélé des activations plus fortes au niveau du gyrus temporal droit (partie moyenne et inférieure), du gyrus fusiforme droit, du putamen droit ainsi que du noyau médiodorsal du thalamus. En HFS, avec un contraste égalisé (EG), les gyri temporal et occipital moyens droits, le gyrus fusiforme droit, une partie de l'OFC droit et gauche ainsi que le gyrus cingulaire postérieur gauche étaient impliqués. La Figure 4.6 présente un aperçu des activations provoquées par le contraste [Apeurée – Neutre] dans les différentes conditions de Fréquences Spatiales et de Contraste. Pour plus de détails, le Tableau C.2, en appendice, présente les coordonnées des pics d'activation obtenus indépendamment des conditions de Fréquences Spatiales et de Contraste ainsi que dans chaque condition expérimentale, associées au nombre de voxels par cluster k, au label AAL, et aux valeurs statistiques T et Z. La Figure C.2, en appendice, présente les cartes des activations obtenues indépendamment des conditions de Fréquences Spatiales et de Contraste. Le contraste inverse [Neutre - Apeurée] n'est pas présenté, car il est considéré comme peu pertinent dans ce travail de thèse.

4.4.2.2 Effet des fréquences spatiales

Le contraste [HFS - BFS], effectué indépendamment des conditions d'EFE (Apeurée, Neutre) et de Contraste (EG, NonEG), implique un réseau plus large que le contraste inverse [BFS - HSF]. Ce réseau comprend notamment les gyri fusiformes droit et gauche (des clusters qui s'étendent sur les gyri occipital et lingual), l'OFC postérieur gauche, le parahippocampe droit, l'insula droite, le gyrus temporal droit (moyen et supérieur) et le CGL droit. De manière intéressante, les HFS activent aussi particulièrement l'amygdale gauche et l'hippocampe droit (un cluster qui s'étend sur l'amygdale droite). Le contraste [BFS-HFS] implique dans l'hémisphère droit le gyrus occipital supérieur et moyen, ainsi que le gyrus temporal moyen, la scissure calcarine et le précuneus. Il implique dans l'hémisphère gauche le gyrus temporal moyen et le gyrus lingual.

Le contraste [HFS - BFS], effectué dans chaque condition d'EFE et de Contraste a révélé les activations suivantes. Avec un visage apeuré, en condition de contraste non égalisé (NonEG), le gyrus lingual droit (un cluster qui s'étendait aux gyri fusiforme et occipital), le gyrus fusiforme gauche (un cluster qui s'étendait aux gyri lingual et occipital), le pulvinar droit, l'insula droite et des régions du lobe temporal droit (supérieur, inférieur) étaient impliqués. Avec un visage apeuré, en condition de contraste égalisé (EG), le contraste [HFS - BFS] a révélé des activations plus élevées dans les gyri fusiformes droit et gauche (des clusters qui s'étendaient aux gyri lingual et occipital), dans le gyrus temporal

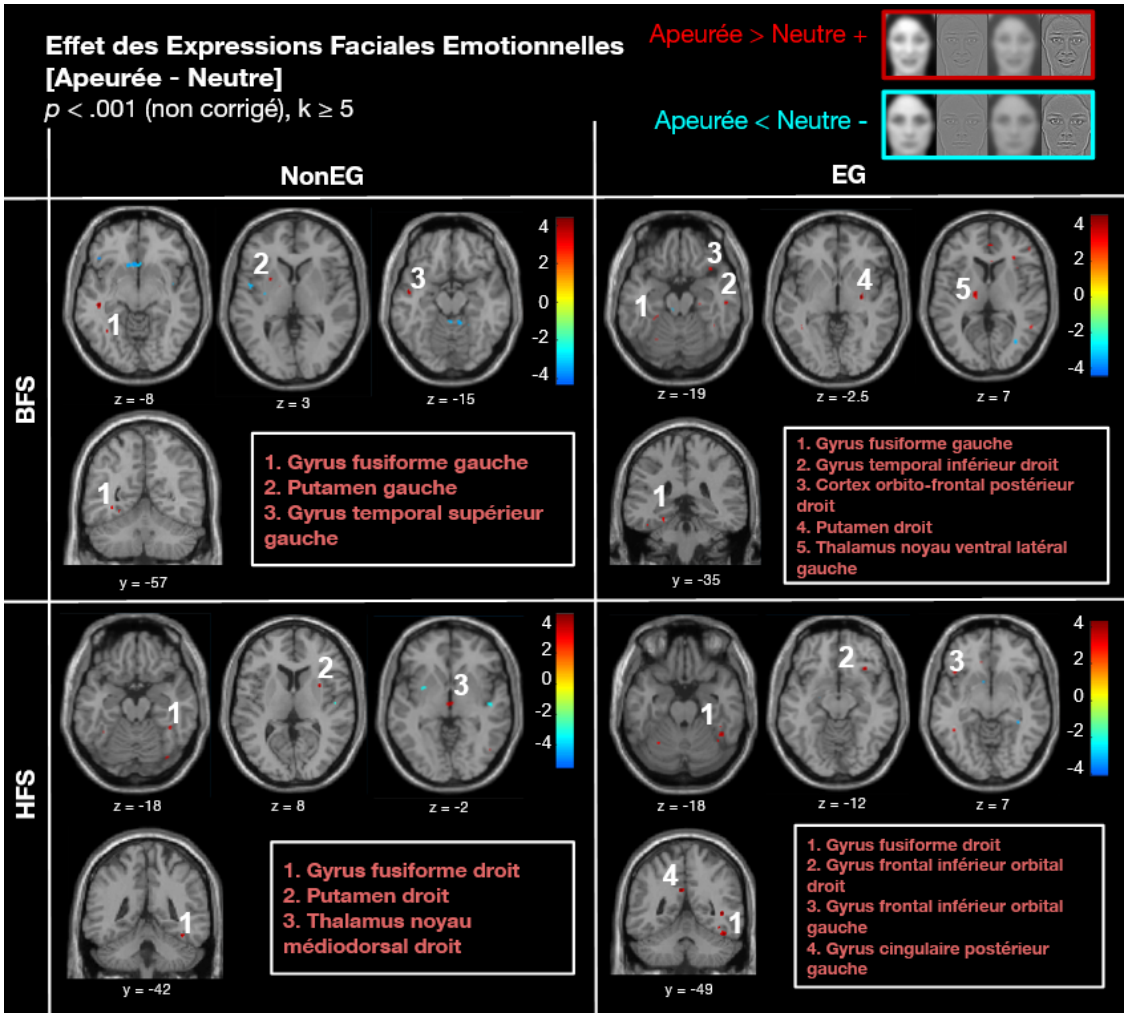

Figure 4.6 – Aperçu des cartes des activations obtenues dans les différentes conditions expérimentales pour le contraste [Apeurée-Neutre], avec un seuil de significativité non corrigé p <.001 et un nombre de voxels par cluster $k \geq 5$. Les valeurs positives, en rouge, correspondent aux régions qui s'activent significativement plus pour les visages apeurés que neutres. Les valeurs négatives, en bleu, correspondent aux régions qui s'activent significativement plus pour les visages neutres que les visages apeurés.

droit (supérieur et moyen) et gauche (supérieur), dans les insulas droite et gauche et le gyrus lingual droit. Avec un visage neutre, en condition de contraste non égalisé (NonEG), il impliquait un réseau qui comprend le gyrus occipital inférieur gauche (qui s'étend aux gyri fusiforme et lingual), le gyrus lingual droit (qui s'étend aux gyri fusiforme et occipital), l'hippocampe droit, le lobe temporal inférieur gauche et la scissure calcarine. Avec un visage neutre, en condition de contraste égalisé (EG), le contraste [HFS - BFS] a révélé des activations plus fortes au niveau des gyri fusiformes droit et gauche (des clusters qui s'étendaient aux gyri occipital et lingual), des amygdales droite et gauche, du putamen droit, de l'OFC droit et du noyau ventral latéral du thalamus gauche.

Le contraste [BFS - HFS], avec un visage apeuré en condition de contraste non égalisé (NonEG), impliquait les gyri temporaux et frontaux moyens droits, le cortex cingulaire antérieur droit et le putamen gauche. Avec un visage apeuré en condition de

contraste égalisé (EG), il a révélé des activations plus prononcées au niveau du précuneus droit et du noyau caudé gauche. Avec un visage neutre, en condition de contraste non égalisé (NonEG), le contraste [BFS - HFS] impliquait le gyrus temporal moyen droit et gauche, le gyrus occipital moyen gauche, l'insula droite et gauche, le noyau médiodorsal du thalamus droit, le gyrus lingual droit et la scissure calcarine droite. Avec un visage neutre, en condition de contraste égalisé (EG), il impliquait l'insula gauche, le gyrus temporal moyen gauche et droit, l'hippocampe droit, le précuneus droit et la scissure calcarine droite. La Figure 4.7 présente un aperçu des activations associées au contraste [BFS – HFS] dans les différentes conditions d'EFE et de Contraste. Les activations négatives correspondent aux régions plus activées par les HFS que les BFS. Pour plus de détails, les Tableaux C.3 et C.4, en appendice, présentent les coordonnées des pics d'activation obtenus, indépendamment des conditions d'EFE et de Contraste ainsi que dans chaque condition, associées au nombre de voxels par cluster k, au label AAL, et aux valeurs statistiques T et Z. La Figure C.3, en appendice, présente les cartes des activations obtenues indépendamment des conditions de Fréquences Spatiales et de Contraste.

4.4.3 Analyse en régions d'intérêt

4.4.3.1 Aire fusiforme des visages (FFA)

L'ANOVA à mesures répétées effectuée sur le PCS moyen dans la FFA (Figure 4.8-a) a révélé un effet principal de l'EFE ($F(1,20) = 17.6$, $p < .001$, $\eta_p^2 = .47$) et des Fréquences Spatiales ($F(1,20) = 38.3$, $p < .001$, $\eta_p^2 = .66$). Le PCS était plus élevé lorsque le visage était apeuré ($M \pm SD$: $1.36 \pm .68$) que neutre ($M \pm SD$: $1.23 \pm .63$), et en HFS ($M \pm SD$: $1.43 \pm .7$) plutôt qu'en BFS ($M \pm SD$: $1.17 \pm .62$).

4.4.3.2 Aire occipitale des visages (OFA)

L'ANOVA à mesures répétées effectuée sur le PCS moyen dans l'OFA (Figure 4.8-b) a révélé un effet principal de l'EFE ($F(1,19) = 13.1$, $p < .001$, $\eta_p^2 = .41$) et des Fréquences Spatiales ($F(1,19) = 40.7$, $p < .001$, $\eta_p^2 = .68$). Le PCS était plus élevé lorsque le visage était apeuré ($M \pm SD$: $1.25 \pm .99$) que neutre ($M \pm SD$: $1.14 \pm .93$), et en HFS ($M \pm SD$: $1.39 \pm .97$) plutôt qu'en BFS ($M \pm SD$: $1 \pm .96$). L'ANOVA a aussi révélé un effet d'interaction entre l'EFE, les Fréquences Spatiales et le Contraste ($F(1,19) = 5.1$, $p = .035$, $\eta_p^2 = .21$). Les comparaisons par paires ont montré que la différence entre les visages apeurés et neutres était seulement significative en HFS quand le contraste n'est pas égalisé ($M \pm SD$ *pour les visages apeurés* : 1.44 ± 1 ; $M \pm SD$ *pour les visages neutres* : 1.26 ± 0.99 ; $p_{corrected} = .011$).

4.4.3.3 Amygdale

Partie latérale : L'ANOVA à mesures répétées effectuée sur le PCS moyen dans la partie latérale de l'amygdale droite (Figure 4.9-a) a révélé un effet principal des Fréquences Spatiales ($F(1,24) = 10.7$, $p = .003$, $\eta_p^2 = .31$) et un effet principal du Contraste ($F(1,24) = 4.26$, $p = .05$, $\eta_p^2 = .15$). Le PCS était plus élevé lorsque le visage était en HFS (M ±

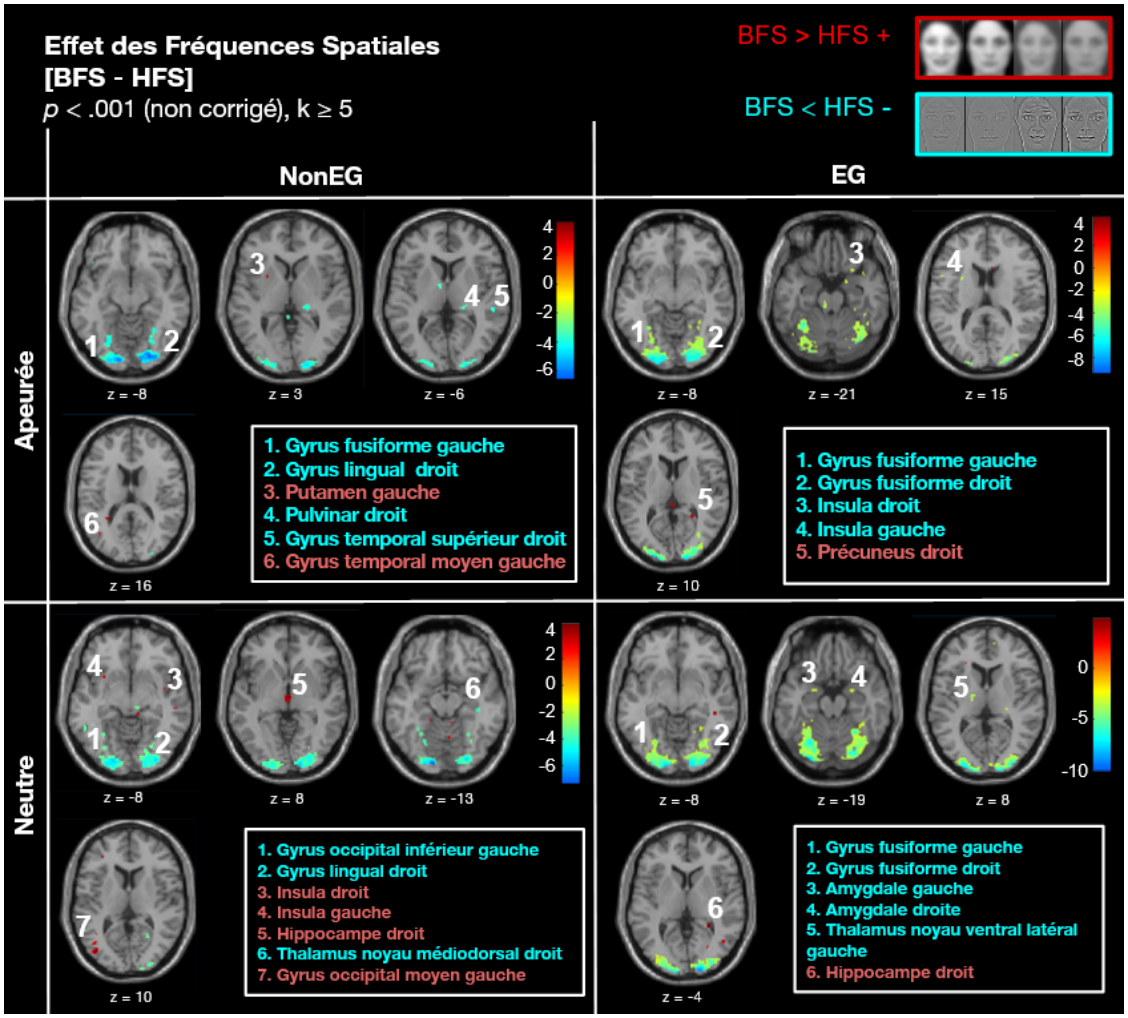

Figure 4.7 – Aperçu des cartes des activations obtenues dans les différentes conditions expérimentales pour le contraste [BFS-HFS], avec un seuil de significativité non corrigé p <.001 et un nombre de voxels par cluster $k \geq 5$. Les valeurs positives, en rouge, correspondent aux régions qui s'activent significativement plus pour les BFS que les HFS. Les valeurs négatives, en bleu, correspondent aux régions qui s'activent significativement plus pour les HFS que les BFS.

SD : 0.07 ± 0.15) plutôt qu'en BFS (M ± SD : -0.017 ± 0.21), et quand le contraste était égalisé ($M \pm SD$: 0.048 ± 0.12) plutôt que non égalisé (M ± SD : 0.005 ± 0.16). Dans la partie latérale de l'amygdale gauche, l'ANOVA a révélé un effet principal des Fréquences Spatiales ($F(1,24) = 8.55$, $p = .007$, $\eta_p^2 = .26$). Le PCS était plus élevé lorsque le visage était en HFS ($M \pm SD$: -0.004 ± 0.21) plutôt qu'en BFS ($M \pm SD$: -0.064 ± 0.2).

Partie médiale : L'ANOVA à mesures répétées effectuée sur le PCS moyen dans la partie médiale de l'amygdale (Figure 4.9-b) a révélé un effet principal des Fréquences Spatiales, aussi bien à gauche qu'à droite (amygdale gauche : $F(1,24) = 6.67$, $p = .016$, $\eta_p^2 = .22$; amygdale droite : $F(1,24) = 17.23$, p <.003, $\eta_p^2 = .42$). Le PCS était plus élevé lorsque le visage était en HFS ($M \pm SD$ *pour l'amygdale gauche :* -0.017 ± 0.17 ; $M \pm SD$ *pour l'amygdale droite* : 0.067 ± 0.20) plutôt qu'en BFS ($M \pm SD$ *pour l'amygdale gauche* : -0.1 ± 0.22 ; $M \pm SD$ *pour l'amygdale droite* : -0.042 ± 0.22).

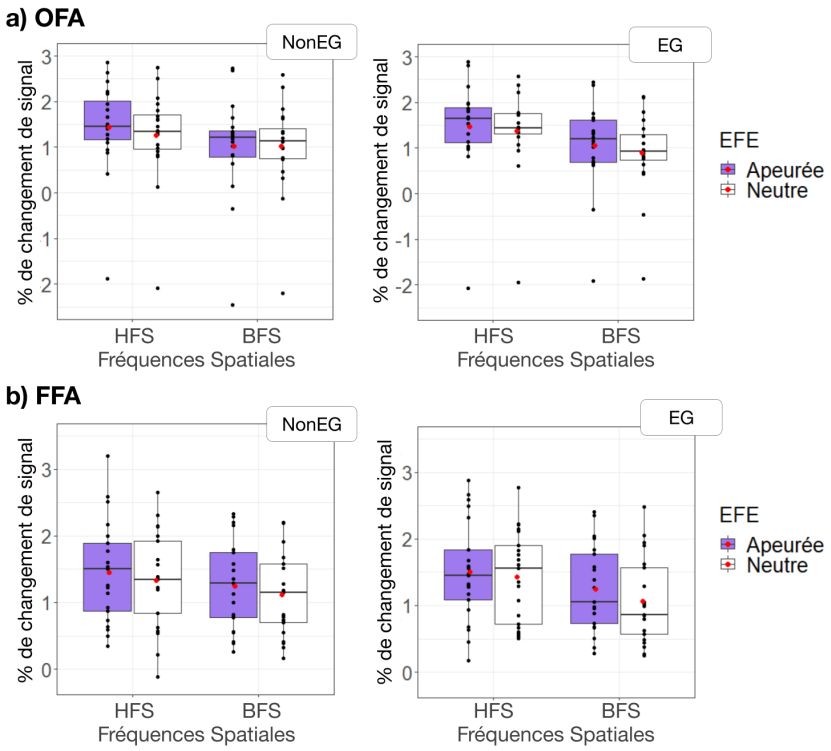

Figure 4.8 – PCS moyen dans (a) l'OFA et (b) la FFA, en condition de contraste non égalisé (à gauche) ou égalisé (à droite), en fonction de l'Expression Faciale Émotionnelle du visage (Apeurée, Neutre) et des Fréquences Spatiales (HFS, BFS).

4.4.3.4 Pulvinar

L'ANOVA à mesures répétées effectuée sur le PCS moyen dans le pulvinar gauche (Figure 4.9-c) n'a révélé aucun effet significatif. L'ANOVA à mesures répétées effectuée sur le PCS moyen dans le pulvinar droit a révélé un effet principal des Fréquences Spatiales ($F(1,24) = 4.9$, $p = .037$, $\eta_p^2 = .01$). Le PCS était plus élevé lorsque le visage était en HFS ($M \pm SD$: -0.094 ± 0.16) plutôt qu'en BFS ($M \pm SD$: -0.14 ± 0.16). L'ANOVA a aussi révélé un effet d'interaction entre les Fréquences Spatiales et le Contraste ($F(1,24) = 13.1$, $p = .001$, $\eta_p^2 = .19$). Les comparaisons par paires ont montré que le PCS était plus élevé pour les HFS que les BFS seulement en condition de contraste non égalisé ($M \pm SD$ *pour les HFS* : -0.65 ± 0.19 ; $M \pm SD$ *pour les BFS* : -0.17 ± 0.18 ; $p_{corrected} = .012$). Aussi, pour les BFS, le PCS était plus élevé avec un contraste égalisé ($M \pm SD = -0.1 \pm 0.16$) que non égalisé ($M \pm SD = -0.17 \pm 0.18$; $p_{corrected} = .016$).

4.4.3.5 Colliculus supérieur (CS)

L'ANOVA à mesures répétées effectuée sur le PCS moyen dans le CS gauche (Figure 4.9-d) n'a révélé aucun effet significatif. En revanche, celle effectuée dans le CS droit a révélé une interaction entre les Fréquences Spatiales et le Contraste ($F(1,24) = 7.86$, $p = .01$, $\eta_p^2 = .25$). Les comparaisons par paires n'ont montré aucune différence

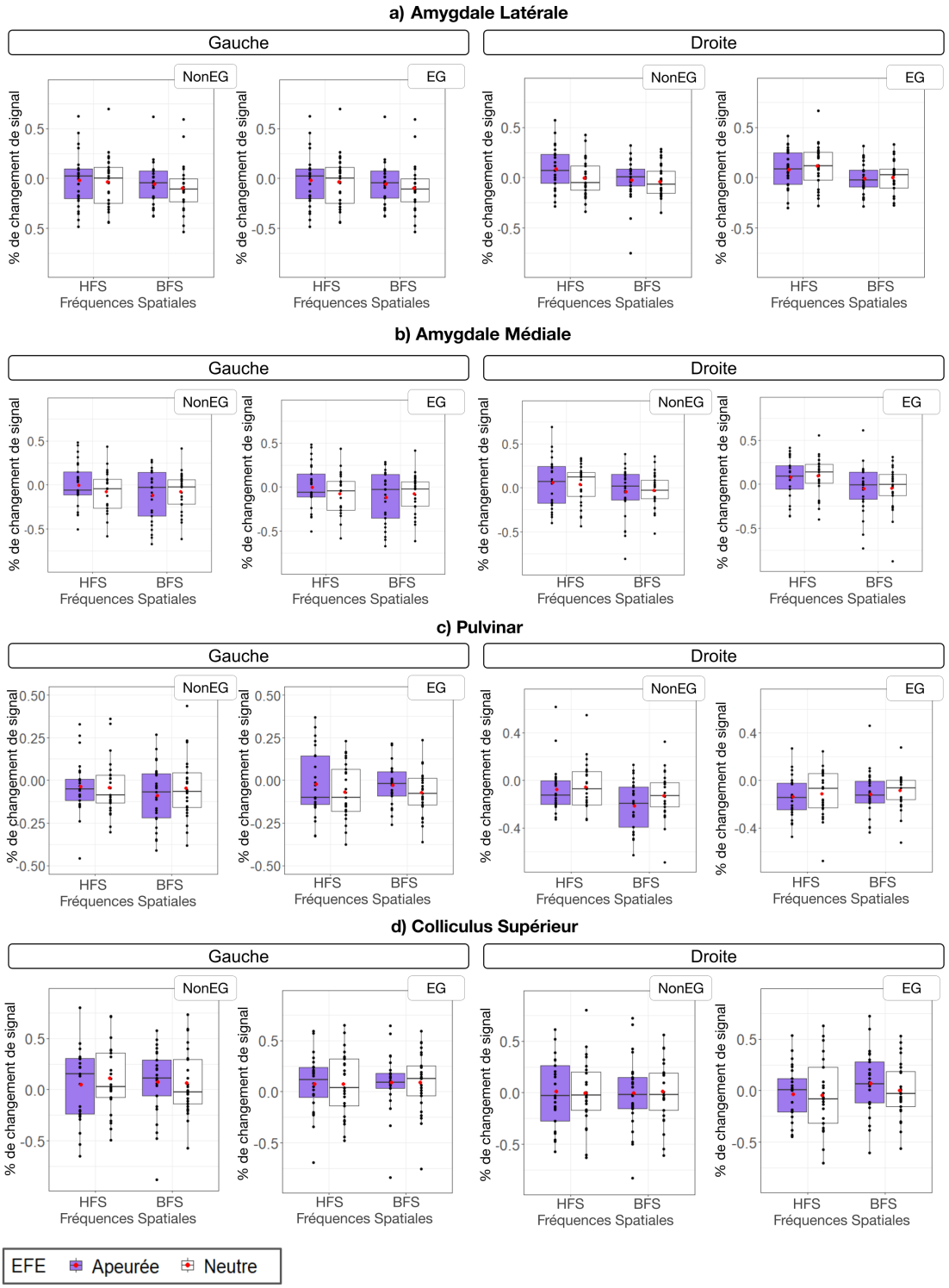

Figure 4.9 – PCS moyen dans (a) l'amygdale latérale, (b) l'amygdale médiale, (c) le pulvinar et (d) le CS, pour les hémisphère gauche et droit, en condition de contraste non égalisé ou égalisé, en fonction de l'Expression Faciale Émotionnelle du visage (Apeurée, Neutre) et des Fréquences Spatiales (HFS, BFS).

significative. Néanmoins, le PCS était marginalement plus élevé en BFS qu'en HSF avec un contraste égalisé ($p_{corrected} = .09$).

4.5 Discussion

Cette expérience avait pour but d'évaluer l'influence des fréquences spatiales et de l'égalisation du contraste sur les réponses neurales de certaines régions impliquées dans le traitement des expressions faciales. Dans un premier temps, nous avons généré les cartes des activations, obtenues sur toute la fenêtre d'acquisition, correspondant à l'effet des fréquences spatiales (HFS ou BFS) et des émotions (visage neutre ou apeuré). Cette première analyse avait pour but de rendre compte des réseaux cérébraux impliqués dans le traitement des émotions et des fréquences spatiales dans notre tâche, et d'évaluer leur concordance avec la littérature. Les cartes des activations correspondant à l'effet des émotions ont mis en avant les régions qui différenciaient les visages neutres des visages apeurés. Nous nous sommes particulièrement intéressés aux régions qui étaient plus activées par les visages apeurés que neutres. Indépendamment des fréquences spatiales et du contraste, ces régions incluaient le gyrus fusiforme (une région qui correspond à la FFA), mais aussi une partie de l'amygdale gauche, conformément à ce qui était attendu (Fusar-Poli et al., 2009). En séparant nos conditions de contraste et de fréquences spatiales, les cartes des activations ont mis en évidence des réseaux partiellement distincts. En effet, seule la FFA s'activait de manière robuste dans chaque condition. Les cartes des activations correspondant à l'effet des fréquences spatiales ont mis en avant un réseau d'activation plus large pour le traitement des HFS que pour le traitement des BFS. Ce réseau incluait des clusters occipito-temporaux très étendus, englobant le gyrus fusiforme, et présents dans chaque hémisphère cérébral. Ces clusters étaient présents dans chaque condition expérimentale, mais étaient encore plus étendus (et l'effet encore plus fort, en termes statistiques) avec un contraste égalisé (c'est-à-dire lorsque le contraste des HFS est rehaussé), comme en témoigne le nombre de voxels par cluster (Tableau C.4). De manière intéressante, les amygdales droite et gauche étaient plus activées pour les HFS que les BFS. En distinguant les conditions d'émotion et de contraste, cet effet n'était significatif qu'avec un contraste égalisé et un visage neutre.

Dans un second temps, nous avons effectué une analyse en régions d'intérêt. Nous avons observé dans les aires sélectives aux visages (OFA et FFA) des activations plus fortes pour les visages apeurés que neutres (pour l'OFA cet effet entrait en interaction avec les fréquences spatiales et le contraste, et les comparaisons par paires ont montré qu'il n'était significatif qu'en HFS, avec un contraste non égalisé), et pour les visages en HFS plutôt qu'en BFS. Pour la FFA, les activations plus fortes pour les visages en HFS qu'en BFS étaient attendues, puisqu'elles sont en accord avec les résultats de l'étude de Vuilleumier et al. (2003). L'effet des émotions observé au niveau de la FFA et de l'OFA a été mis en évidence dans plusieurs études utilisant des images non filtrées (C. J. Fox et al., 2009 ; Ganel et al., 2005 ; Kadosh et al., 2011 ; Liu et al., 2021 ; Vuilleumier et Pourtois, 2007 ; Xu et al., 2021). Dans l'étude de Vuilleumier et al., cet effet n'était observé pour

la FFA qu'avec des visages en BFS et non filtrés. Dans notre étude, il est aussi observé avec des visages en HFS, ce qui suggère que cet effet n'est pas dépendant du contenu fréquentiel. L'analyse en région d'intérêt effectuée dans les régions correspondant à la voie sous-corticale n'a pas révélé les effets attendus. L'activité observée au niveau de l'amygdale, aussi bien à droite qu'à gauche (et aussi bien dans la partie médiale que latérale) était plus forte pour les HFS que les BFS. Bien qu'une petite partie de l'amygdale était sensible aux émotions (6 voxels, comme en témoigne l'analyse des cartes des activations), nous n'avons observé aucun effet significatif des émotions dans notre analyse en région d'intérêt, ni sous la forme d'un effet principal ni sous la forme d'une interaction avec les autres variables. Ainsi, ces résultats ne vont pas dans le sens de notre hypothèse concernant un effet des émotions porté par les BFS. Des activations plus fortes pour les HFS que les BFS ont aussi été observées au niveau du pulvinar droit, un effet néanmoins seulement significatif en condition de contraste non égalisé. Au niveau du CS, les activations étaient marginalement plus fortes en BFS qu'en HFS en condition de contraste non égalisé.

Ainsi, nous n'avons pas observé d'effet des expressions faciales dans les régions correspondant à la voie sous-corticale, notamment l'amygdale, qui est pourtant connue pour répondre plus fortement aux visages apeurés (Fusar-Poli et al., 2009). Ces résultats rejoignent ceux d'autres études utilisant des images filtrées qui n'ont pas observé d'effet des émotions en BFS dans l'amygdale (Corradi-Dell'Acqua et al., 2014 ; Ottaviani et al., 2012). De plus, notre étude a mis en avant des activations plus fortes pour les HFS que les BFS dans de nombreuses régions, un résultat que nous n'avions pas attendu. Dans l'étude de Vuilleumier et al. (2003) des activations plus prononcées pour les HFS que les BFS avaient également été observées au niveau des gyri fusiformes, temporaux et occipitaux. Cependant, les clusters d'activations n'étaient pas aussi étendus que dans notre étude. Aussi, l'effet que les auteurs avaient observé était plus fort dans l'hémisphère gauche que dans l'hémisphère droit, ce qui n'est pas systématiquement le cas notre étude.

Il est possible qu'un biais dans la création de nos stimuli nous empêche d'observer l'effet des émotions auquel nous nous serions attendus, et avantage le traitement des HFS. Dans notre étude, nous avons utilisé un filtre gaussien pour obtenir nos images filtrées. Plus précisément, deux filtres gaussiens ont été créés, l'un laissant passer les HFS et l'autre les BFS, et ont été multipliés au spectre d'amplitude des images. Cette méthode a pour avantage d'éviter la création d'artefacts apparaissant avec un filtrage strict, dit *Heaviside*. Une alternative aux filtres gaussiens et *Heaviside* est le filtre *Butterworth*, qui produit des images plus fines que le filtrage gaussien, avec moins d'artefacts qu'un filtrage *Heaviside* (Dogra et Bhalla, 2014 ; Makandar et Halalli, 2015 ; Perfetto et al., 2020). Dans une étude de Perffetto et al. (2020), il a été montré que la forme du filtre (gaussien, *Heaviside* ou *Butterworth*) peut influencer les performances dans une tâche de catégorisation de scènes. Dans leur étude, les participants devaient catégoriser des scènes visuelles selon les catégories suivantes : plage, forêt, montagne, autoroute, ville ou bureau. Les performances étaient meilleures en BFS qu'en HFS avec un filtre *Heaviside*, alors qu'avec un filtre gaussien l'effet inverse était observé (aucune différence significative n'était observée avec un filtre *Butterworth*). Ainsi, l'utilisation d'un filtre gaussien pourrait

favoriser la reconnaissance d'images en HFS. Dans l'étude de Vuilleumier et al. (2003) la forme du filtre utilisé n'est pas précisée. Visuellement, leurs stimuli semblent correspondre à des bandes de fréquences plus hautes que celles que nous avons, bien que les fréquences de coupures utilisées dans leur étude et la nôtre soient les mêmes (Figure 4.10).

En observant nos stimuli, nous pouvons remarquer que les yeux (les détails, comme le contour blanc autour de l'iris) sont très visibles en HFS, et presque indiscernables en BFS. Or, plusieurs études suggèrent qu'il existe un lien fort entre le décodage de l'information transmise par les yeux et l'activité de l'amygdale (Adolphs et al., 2005 ; Ahs et al., 2014 ; Asghar et al., 2008 ; Kennedy et Adolphs, 2010 ; M. J. Kim et al., 2016 ; Meletti et al., 2012 ; Morris et al., 2002 ; Whalen et al., 2004). Par exemple, une étude de Ahs et al. (2014) a montré à l'aide d'une analyse en composante principale que les deux composantes qui influencent le plus l'activité de l'amygdale sont les déplacements de sourcils et la proportion de blanc dans les yeux. Ces résultats suggèrent que la représentation des expressions dans l'amygdale serait basée sur des différences physiques (portées par la région des yeux) plutôt que sur des différences en termes de catégories d'émotions. Ces résultats sont en accord avec ceux d'autres études qui montrent une sensibilité de l'amygdale au blanc des yeux (M. J. Kim et al., 2016 ; Whalen et al., 2004), à la taille de la pupille (Demos et al., 2008) ou à l'orientation du regard (Adams Jr et al., 2003). Cette hypothèse d'un lien entre l'amygdale et le traitement des yeux est aussi corroborée par des études chez un patient atteint d'une lésion bilatérale de l'amygdale. En effet, ces études ont montré une exploration anormale des visages lors du jugement d'expressions faciales, qui se traduit par une diminution du nombre de fixations sur la région des yeux (Adolphs et al., 2005 ; Kennedy et Adolphs, 2010). Ces données suggèrent que l'amygdale pourrait utiliser les yeux écarquillés comme une approximation grossière de la présence de visages apeurés, et offrent une explication perceptive à un mécanisme impliquant le jugement de stimuli sociaux complexes (Whalen et al., 2013). L'amygdale est également plus sensible aux objets pointus qu'aux objets courbés (Bar et Neta, 2007), et dans notre étude les BFS offrent plus de courbes que les HFS, qui ont des contours plus fins. Ainsi, nous pouvons supposer que l'absence d'effet des émotions en BFS, ainsi que les réponses moins prononcées en BFS, viennent du fait que les yeux sont difficiles à distinguer.

Une autre explication à l'avantage de HFS concerne le temps de présentation de nos images. Par exemple, une étude réalisée par Goffaux et al. (2011) suggère que le temps de présentation des stimuli influence le traitement des fréquences spatiales dans les régions sélectives aux visages. Dans cette étude, les auteurs ont observé une réponse de la FFA plus forte pour les BFS que les HFS lorsque les stimuli étaient présentés 75 ms. Mais, lorsque les stimuli étaient présentés plus longtemps (à partir de 150 ms), c'est l'effet inverse qui était observé. Ainsi, le fait que dans notre étude les images soient présentées pendant une durée relativement longue (200 ms) pourrait favoriser le traitement des HFS.

Au final, nos résultats remettent en question l'implication de la voie sous-corticale dans le traitement des expressions faciales. Comme nous l'avons évoqué dans les paragraphes précédents, cela pourrait être dû à un biais méthodologique induit par le filtrage spatial. Cependant, ce résultat rejoint certains travaux de la littérature. Par exemple,

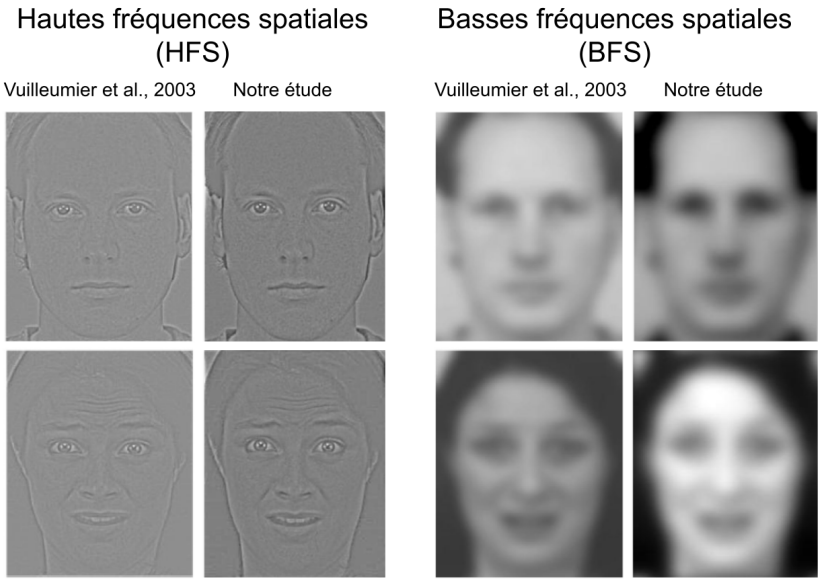

Figure 4.10 – Comparaison des stimuli utilisés dans notre étude et celle de Vuilleumier et al. (2003), en HFS (droite) et en BFS (gauche). Nous avons volontairement affiché des images qui étaient, avant toute manipulation, identiques dans les deux études. Bien qu'une partie des images utilisées dans les deux études était issue de la même base de données, ce n'était pas le cas de tous les stimuli. Pour notre étude, ce sont les stimuli en condition de contraste non égalisé qui sont affichés puisque, dans leur étude, les stimuli ne sont pas égalisés en contraste après le filtrage.

les résultats des études de McFadyen et al. (2017) et Garvert et al. (2014) cités en introduction. Pour rappel, ces études ont utilisé la DCM pour évaluer la connectivité entre les régions lors de la perception de visages exprimant différentes émotions à partir d'un enregistrement de l'activité en MEG. D'après leurs résultats, les modèles incluant la voie sous-corticale sont les plus probables. Néanmoins, cette voie ne serait pas influencée par les émotions du visage. De plus, certains modèles du traitement des visages suggèrent que la voie sous-corticale serait impliquée de manière plus globale dans la détection des visages (Johnson, 2005 ; Johnson et al., 2015). Pour Johnson et al. (2015), cette voie pourrait être à l'origine de résultats d'études chez les nourrissons qui montrent que certaines informations sur les caractéristiques des visages sont disponibles dès la naissance. En effet, les nouveau-nés humains s'orientent préférentiellement vers des modèles schématiques simples ressemblant à des visages (voir par exemple Johnson et al., 1991). Ce modèle suggère que les activations plus prononcées pour les expressions apeurées pourraient venir du fait qu'ils constituent un modèle de visage pertinent pour la survie. Finalement, certaines études chez des patients ayant subi une lésion de l'amygdale n'ont pas observé de détérioration des réponses émotionnelles, ce qui suggère que l'amygdale n'est pas indispensable pour traiter des stimuli émotionnels (Bach et al., 2015 ; Piech et al., 2011 ; Piech et al., 2010 ; Tsuchiya et al., 2009 ; Wang et al., 2014).

Nous pouvons aussi noter certaines différences concernant les méthodes d'analyse statistique entre notre étude et celle de Vuilleumier et al. (2003), qui pourraient expliquer

les différences en termes de résultats obtenus. Par exemple, la plupart des clusters correspondant à l'amygdale ont été obtenus dans leur étude avec une valeur statistique $p > .001$ en non corrigé (plus précisément, $p < .005$ ou $p < .01$). En IRMf, un seuil de valeur $p < .001$ non corrigé pour une analyse dans la fenêtre d'acquisition entière est déjà critiqué en considérant la quantité de voxels analysés. Ces valeurs augmentent donc d'autant plus le risque d'obtenir des résultats faussement positifs et peuvent ainsi rendre la réplication compliquée (Lieberman et Cunningham, 2009).

Pour finir, il est intéressant de noter qu'au niveau du CS droit nous avons obtenu une interaction entre les fréquences spatiales et le contraste, guidée par une activation marginalement plus forte pour les BFS que les HFS en condition de contraste non égalisé. Bien que la différence soit marginale après la correction appliquée sur la valeur p, elle peut être liée à certaines études sur le profil des réponses des cellules du CS. Par exemple, des données issues d'enregistrements intracrâniens chez des rongeurs montrent que, dans certains neurones, la réponse augmente de manière relativement linéaire avec l'augmentation du contraste. Mais, dans de nombreux neurones, la réponse augmente jusqu'à atteindre un pic à des contrastes intermédiaires et ne plus augmenter avec la hausse du contraste. Dans d'autres neurones encore, la réponse atteint un pic à un contraste intermédiaire et diminue avec l'augmentation du contraste (De Franceschi et Solomon, 2020). Chez les humains, des données IRMf ont mis en évidence une activation du CS dès 5% de contraste ainsi qu'une augmentation de son activité pour un contraste de 10% et une saturation ensuite (Schneider et Kastner, 2005). Ces données suggèrent que le CS ne serait sensible aux changements de contraste que pour des valeurs de contraste faibles. Plusieurs études ont aussi souligné une sensibilité du CS spécifique aux BFS. Par exemple, il ne répondrait plus aux changements de fréquences spatiales pour des fréquences supérieures à 2 cpd chez le chat (Bisti et Sireteanu, 1976) et 4 cpd chez le singe (C.-Y. Chen et al., 2018). Et, en moyenne, il serait plus actif pour des fréquences spatiales se situant aux alentours de 0.08 cpd (Mimeault et al., 2004). Ainsi, bien que l'effet d'interaction observé sur le CS n'était pas attendu (car nous avons basé nos hypothèses sur l'étude de Vuilleumier et al., 2003), il est accord avec les études citées précédemment, qui soulignent une préférence du CS pour les BFS et une faible sensibilité à des contrastes élevés.

Chapitre 4 - Points clés

- Mise en place d'une expérience en IRMf.
- Des réponses plus fortes face à des visages apeurés que neutres ont été observées au niveau de la FFA, de l'OFA, du gyrus temporal, de l'insula gauche, d'une petite partie de l'amygdale gauche et de l'OFC droit.
- Des réponses plus fortes face à des visages en HFS qu'en BFS ont été observées dans de larges clusters occipito-temporaux incluant le gyrus fusiforme, la FFA et l'OFA, ainsi qu'au niveau de l'amygdale et du pulvinar.
- L'analyse en région d'intérêt effectuée sur l'amygdale (ainsi que sur le pulvinar et le CS) n'a pas montré de différences d'activations en fonction des émotions, ni d'interaction avec les fréquences spatiales et le contraste.
- Les activations plus élevées pour les HFS que les BFS dans l'amygdale pourraient s'expliquer par les caractéristiques de nos stimuli. Par exemple, en BFS, les yeux, qui ont souvent été mis en lien avec l'activité de l'amygdale, sont difficiles à distinguer.
- L'égalisation du contraste est associée à un élargissement des clusters d'activations occipito-temporaux qui montraient des réponses plus fortes pour les HFS que les BFS.

Chapitre 5
Discussion générale

Table des matières

5.1 Préface

L'objectif de ce travail de thèse était d'étudier l'influence des expressions faciales dans la programmation des saccades oculaires. Comme nous l'avons vu dans les chapitres précédents, la programmation des saccades a lieu dans une carte de priorité, située dans les couches intermédiaires du CS. Cette carte de priorité est un espace en deux dimensions, suivant une organisation rétinotopique, dans lequel chaque localisation de la scène possède un poids. Ce poids correspond à une valeur de priorité qui est influencée

à la fois par des facteurs *bottom-up* et par des facteurs *top-down*. L'information issue des facteurs *bottom-up* est transmise à la carte de priorité depuis les cartes de saillance. L'information concernant la pertinence émotionnelle, qui supporterait un traitement privilégié des visages émotionnels, est transmise par l'amygdale. Néanmoins, les processus par lesquels opère cette transmission sont encore incertains.

Une possibilité repose sur l'existence d'une voie sous-corticale, reliant le CS à l'amygdale en passant par le pulvinar. Les études chez des patients atteints de cécité corticale montrent que ces patients sont capables de répondre inconsciemment aux expressions faciales de visages présentés dans leur zone aveugle (Celeghin et al., 2015), suggérant l'existence d'un traitement indépendant du cortex visuel. La particularité de cette voie sous-corticale est qu'elle traiterait uniquement l'information grossière, en BFS. De plus, elle serait particulièrement impliquée dans la discrimination des expressions faciales. Ainsi, certaines études ont mis en évidence des activations de l'amygdale plus fortes pour les visages apeurés que neutres, mais seulement pour des images qui contenaient des BFS (Méndez-Bértolo et al., 2016 ; Vuilleumier et al., 2003). Cette voie opérerait en parallèle à la voie corticale (qui part de la rétine et se projette vers le cortex visuel primaire via un relais dans le CGL), qui aurait quant à elle accès à l'information détaillée, en HFS. Bien que l'hypothèse de l'existence de la voie sous-corticale ait été étudiée de manière extensive, elle fait encore l'objet de nombreux débats (de Gelder et al., 2011 ; Pessoa, 2010).

Dans ce travail de thèse, nous avons testé l'hypothèse d'une modulation rapide (< 100 ms) des mécanismes de programmation des saccades oculaires, qui favoriserait l'orientation du regard vers des visages émotionnels, particulièrement apeurés. Nous supposions que cet effet soit indépendant de l'objectif de l'observateur et originaire du traitement des BFS au sein de la voie sous-corticale. Nous avons mené plusieurs expériences comportementales, faisant intervenir un paradigme de choix saccadique et des visages avec une expression neutre, joyeuse ou apeurée. Afin de mieux comprendre certains résultats comportementaux obtenus au cours de ce travail, nous avons utilisé des modèles computationnels. Plus précisément, nous avons utilisé des réseaux de neurones artificiels, qui nous ont permis de mieux caractériser les processus impliqués dans la détection d'expressions faciales émotionnelles. Pour finir, nous avons également mené une expérience en IRMf dans le but d'évaluer les bases cérébrales du traitement des expressions faciales, en fonction des fréquences spatiales et de l'égalisation du contraste.

Dans ce chapitre, nous commencerons par rappeler brièvement les résultats obtenus. Puis, nous discuterons des apports théoriques et méthodologiques de ce travail de thèse. Enfin, nous présenterons certaines de ses limites, ainsi que des perspectives envisagées.

5.2 Synthèse des résultats obtenus

Dans l'**Expérience 1**, nous avons utilisé un paradigme de choix saccadique opposant un véhicule à un visage neutre, joyeux ou apeuré. Les participants devaient faire une saccade vers le véhicule dans une session, et vers le visage dans une autre session. Nous avons observé que les saccades vers les visages étaient effectuées plus rapidement que les

saccades vers les véhicules, et elles étaient aussi plus souvent dans la bonne direction. Néanmoins, ces effets n'étaient pas affectés par les expressions faciales du visage. Les saccades atterrissaient plus bas dans le visage lorsqu'il était joyeux que lorsqu'il était apeuré ou neutre, et lorsqu'il était apeuré que lorsqu'il était neutre.

Ensuite, dans l'**Expérience 2**, nous avons utilisé un paradigme de choix saccadique opposant un visage neutre à un visage émotionnel, joyeux ou apeuré. Les participants devaient faire une saccade vers le visage neutre dans une session, et vers le visage émotionnel dans une autre session. Nous avons observé que les saccades vers les visages émotionnels étaient effectuées plus rapidement que les saccades vers les visages neutres. Elles étaient aussi plus souvent dans la bonne direction, particulièrement lorsque le visage émotionnel était joyeux. Les saccades atterrissaient plus bas dans le visage lorsqu'il était joyeux que lorsqu'il était apeuré.

La **simulation de l'Expérience 2** nous a permis de quantifier les différences physiques entre les visages neutres et joyeux et entre les visages neutres et apeurés. Nous avons utilisé un réseau de neurones artificiel (ici, un MLP) testé et entraîné sur sa capacité à distinguer des paires de visages en fonction de la position du visage neutre et du visage émotionnel dans la paire (en d'autres termes, selon que la paire était de type émotionnel-neutre ou de type neutre-émotionnel). Les résultats ont montré que le réseau de neurones artificiel réussissait mieux la tâche avec un visage joyeux plutôt qu'avec un visage apeuré.

Dans l'**Expérience 3**, nous avons utilisé un paradigme de choix saccadique opposant un visage féminin à un visage masculin. L'un des visages était neutre, l'autre émotionnel (joyeux ou apeuré), et les participants devaient faire une saccade vers le visage masculin dans une session, et vers le visage féminin dans une autre session. Nous avons observé que les saccades vers les visages masculins ou féminins étaient plus souvent dans la bonne direction lorsque le visage cible était émotionnel, d'autant plus joyeux, que lorsqu'il était neutre. Les saccades atterrissaient plus bas dans le visage lorsqu'il était joyeux plutôt que lorsqu'il était apeuré ou neutre, et lorsqu'il était apeuré que lorsqu'il était neutre.

Puis, dans l'**Expérience 4**, nous avons reproduit un paradigme de choix saccadique opposant un visage neutre à un visage émotionnel, joyeux ou apeuré. Les participants devaient faire une saccade vers le visage neutre dans une session, et vers le visage émotionnel dans une autre session. Mais, cette fois, les images étaient présentées sous différentes conditions de filtrages : en HFS, en BFS ou non filtrées. De plus, un groupe de participants a passé l'expérience avec un contraste égalisé entre les fréquences, et un autre sans égalisation du contraste entre les fréquences. Nous avons observé que les saccades vers les visages émotionnels étaient effectuées plus rapidement que les saccades vers les visages neutres. Elles étaient aussi plus souvent dans la bonne direction, particulièrement lorsque le visage émotionnel était joyeux. Les saccades étaient plus souvent dans la bonne direction lorsque les images étaient en HFS qu'en BFS. Elles étaient aussi effectuées plus rapidement, un effet seulement significatif lorsque le contraste était égalisé. Les saccades atterrissaient plus bas dans le visage lorsqu'il était émotionnel que lorsqu'il était neutre, et lorsqu'il était émotionnel, elles atterrissaient plus bas lorsqu'il était joyeux que lorsqu'il était apeuré.

L'analyse de saillance effectuée sur les stimuli de l'**Expérience 4** a permis de mettre en évidence les zones qui se distinguent par leurs caractéristiques physiques. De plus, nous avons montré que la saillance de la région de la bouche, contrairement à la saillance de la région des yeux, corrélait avec les performances des participants dans l'**Expérience 4**.

La **simulation de l'Expérience 4** nous a permis de générer des cartes de saillance spécifiques à une tâche. Ainsi, nous avons mis en évidence les régions de nos stimuli particulièrement utiles dans une tâche similaire à celle de l'**Expérience 4**. Comme pour la **simulation de l'Expérience 2**, nous avons utilisé un réseau de neurones artificiel (ici, un CNN) testé et entraîné sur sa capacité à distinguer des paires de visages en fonction de la position du visage neutre et du visage émotionnel dans la paire. Les résultats ont mis en avant une région particulièrement utile : la bouche. Nous avons observé que la saillance de la bouche, qu'elle soit extraite depuis les cartes de saillance *bottom-up* ou depuis les cartes du CNN prédisait de manière significative les performances des participants dans l'**Expérience 4**. Les prédictions étaient néanmoins légèrement meilleures en se basant sur les cartes du CNN.

Pour finir, dans l'**Expérience 5**, nous avons étudié l'activité en IRMf lors d'une tâche de catégorisation du genre de visages neutres ou apeurés. Les visages étaient présentés en HFS ou en BFS, avec un contraste égalisé ou non. Nous avons observé des réponses plus fortes face à des visages apeurés que neutres au niveau de la FFA et de l'OFA. Les HFS étaient associées à des réponses plus fortes au niveau de larges clusters occipito-temporaux, incluant le gyrus fusiforme, la FFA et l'OFA, ainsi qu'au niveau de l'amygdale et du pulvinar droit. L'égalisation du contraste était associée à un élargissement des clusters occipito-temporaux qui répondaient plus fortement aux visages en HFS qu'aux visages en BFS. Contrairement à nos hypothèses, l'analyse en région d'intérêt effectuée dans l'amygdale (ainsi que dans le pulvinar et le CS) n'a pas montré de sensibilité aux expressions faciales, que ce soit sous la forme d'un effet principal ou sous la forme d'interactions avec les fréquences spatiales ou le contraste.

5.3 Apports théoriques

Dans cette section, nous allons revenir sur les implications théoriques de nos résultats. Nous discuterons d'abord de l'aspect automatique de la capture du regard par les visages émotionnels. Ensuite, nous reviendrons sur la détection des visages émotionnels en abordant les questions suivantes : comment sont utilisées les fréquences spatiales ? Quelles sont les régions utiles ? Pour finir, nous discuterons des bases cérébrales de la perception des expressions faciales après avoir abordé l'influence des expressions faciales sur les points d'arrivée des saccades.

5.3.1 Une capture automatique du regard par les visages émotionnels ?

L'une des hypothèses de ce travail de thèse suggère que les visages émotionnels peuvent capturer l'attention et le regard plus efficacement que les visages neutres d'une manière automatique (c'est-à-dire rapidement et indépendamment de l'attention). Dans

l'**Expérience 1**, la présence d'une expression faciale émotionnelle sur un visage n'a pas facilité sa détection et le déploiement d'une saccade vers lui. Ainsi, les visages étaient détectés et fixés plus efficacement que les véhicules, mais les expressions faciales n'avaient pas d'influence sur cet effet. Du fait qu'il est difficile de conclure sur une absence d'effet (qui peut s'expliquer par un manque de puissance statistique), nous avons calculé les *Bayes factors* correspondant à l'hypothèse nulle dans notre étude. Ici, l'hypothèse nulle suggère qu'il n'y a pas de différence entre les latences ou les proportions de saccades correctes associées aux différentes expressions faciales. Les indices obtenus traduisent une évidence très forte en faveur de l'hypothèse nulle (Jeffreys, 1989), ce qui rend notre interprétation plus légitime. Ainsi, nos résultats ne soutiennent pas l'idée d'une capture automatique du regard par les visages émotionnels. Ce n'est pas la première fois que de tels résultats sont mis en avant. Par exemple, des résultats similaires (c'est-à-dire une capture rapide du regard par les visages, mais qui ne favorise pas les visages émotionnels) avaient été observés dans l'étude de Devue et Grimshaw (2017). Pour rappel, les auteurs avaient testé l'effet des expressions faciales dans une tâche où les visages étaient connus pour attirer particulièrement le regard (Devue et al., 2012). Un cercle de points colorés était affiché à l'écran et les participants devaient effectuer une saccade vers un point d'une certaine couleur. Des images de différents objets non pertinents, dont un visage neutre ou en colère, étaient affichées dans un cercle concentrique à l'intérieur du cercle de points. Les auteurs ont constaté que les visages attiraient davantage le regard que les autres objets, mais cette attraction n'était pas plus forte avec un visage en colère qu'avec un visage neutre. Dans une étude en EEG, Kulke (2019) a observé que les latences de saccades dirigées vers des visages présentés en périphérie ne différaient pas en fonction de l'expression faciale. Ces résultats peuvent être liés à une hypothèse avancée dans la revue de la littérature de Mulckhuyse (2018). Cette hypothèse suggérait qu'une capture de l'attention par les visages émotionnels n'apparaîtrait que tardivement dans le décours temporel de l'analyse visuelle, environ 200 ms après l'apparition des stimuli. Or, dans les études précédemment citées ainsi que dans l'**Expérience 1**, les saccades sont effectuées très rapidement, souvent en moins de 200 ms. Dans notre Expérience 1 par exemple, elle étaient effectuées de manière fiable dès 110 ms, et en moyenne en 176 ms. Dans les autres études citées elles étaient effectuées en moyenne en 188 ms (Kulke, 2019) et en 199 ms (Devue et Grimshaw, 2017).

Ensuite, nos résultats ont montré que, dans d'autres tâches, les visages émotionnels sont privilégiés par rapport aux visages neutres. Que ce soit dans le cadre de la discrimination d'émotions (**Expériences 2 et 4**) ou dans le cadre de la discrimination de genre (**Expérience 3**). Nous avons observé de manière constante au cours de ces trois expériences qu'ils sont plus souvent la cible de la première saccade. Dans les **Expériences 2 et 4**, nous avons aussi observé qu'ils étaient fixés plus rapidement que les visages neutres. Ainsi, ces résultats nous montrent que la présence d'une expression faciale émotionnelle sur un visage peut faciliter son traitement et la programmation d'une saccade vers lui. Néanmoins, un doute subsiste quant à savoir si ce processus est soutenu par l'interprétation du contenu émotionnel ou l'analyse de la configuration physique. Dans l'**Expérience 2**, il est possible que les participants choisissent de vérifier

d'abord les visages émotionnels, même dans le cadre de la recherche du visage neutre, en recherchant une bouche ouverte, qui est une caractéristique particulièrement saillante (Calvo et Lundqvist, 2008 ; Horstmann et al., 2012 ; Stuit et al., 2021). Une telle stratégie donnerait la priorité aux visages émotionnels, non pas parce qu'ils possèdent un contenu émotionnel pertinent, mais parce qu'ils ont des configurations qui les rendent plus faciles à détecter. En appui à cette idée, une étude sur la détection et la catégorisation d'émotions suggère que des dents visibles sont particulièrement utiles (Sweeny et al., 2013). Dans l'**Expérience 3**, les meilleures performances obtenues dans le cadre de la détection du genre pourraient également être expliquées par la saillance des visages émotionnels. Nous pouvons supposer que certaines caractéristiques saillantes, comme la bouche des visages émotionnels, favorisent l'orientation de l'attention vers elles.

Dans les **Expériences 2 et 4**, les émotions du visage étaient pertinentes pour la tâche, ce qui n'était pas le cas dans les **Expériences 1 et 3**. Ces résultats témoignent d'une capture de l'attention et du regard par les visages émotionnels qui interviendrait aussi bien dans le cadre d'un traitement implicite que d'un traitement explicite des expressions faciales. Plusieurs études avaient déjà observé un effet des expressions faciales dans des tâches dans lesquelles elles n'étaient pas pertinentes. Par exemple, des tâches dans lesquelles un visage émotionnel est brièvement présenté en amorce, suivi d'un mot ou d'une scène visuelle à catégoriser. La tâche des participants consiste à ignorer le visage (qui n'est donc pas pertinent pour la tâche) et à juger le mot ou la scène comme étant agréable ou désagréable. Classiquement, les temps de réaction sont plus rapides lorsque l'amorce et le stimulus à évaluer sont congruents d'un point de vue affectif, par exemple pour des mots positifs suivant un visage joyeux (Aguado et al., 2007 ; Lipp et al., 2009 ; McLellan et al., 2010 ; Sassi et al., 2014). Dans le cadre de la programmation de saccade, D'Hondt et al. (2016) ont utilisé un paradigme de choix saccadique avec des paires de scènes, l'une ovale et l'autre rectangulaire, présentées à différentes excentricités. Les auteurs ont mis en évidence de meilleures performances lorsque la cible était une scène émotionnelle qu'une scène neutre (cet effet était cependant limité à une excentricité de 10°). Ainsi, nous pouvons supposer que la capture de l'attention et du regard par les expressions faciales n'est pas nécessairement dépendante de leur pertinence pour la tâche.

En résumé, nos résultats ne supportent pas l'idée que les visages émotionnels capturent automatiquement le regard. Nous pouvons supposer que le système visuel, ou les caractéristiques du visage elles-mêmes ont évolué de manière à ce que les visages, qui sont plus susceptibles d'être pertinents que d'autres objets, puissent être détectés très rapidement sur la base de caractéristiques de bas niveau (Baron-Cohen, 1995 ; Emery, 2000 ; Haxby et al., 2000 ; Kobayashi et Kohshima, 1997 ; Lacruz et al., 2019 ; Leopold et Rhodes, 2010 ; Wu et al., 2014). Que cette détection soit basée sur des caractéristiques isolées (par exemple les yeux ; Kauffmann et al., 2021 ; Lewis et Edmonds, 2003) ou sur des informations issues du spectre d'amplitude (Crouzet et Thorpe, 2011 ; Honey et al., 2008), elle pourrait être insuffisante pour décoder les expressions faciales. L'objectif de ce traitement rapide des visages pourrait donc être de placer le visage en vision centrale, pour ensuite procéder à une analyse plus fine. Cette analyse plus fine permettrait d'extraire davantage de

caractéristiques, telles que l'expression émotionnelle. Néanmoins, nos résultats supportent l'idée que les visages émotionnels peuvent capturer le regard indépendamment de la tâche de l'observateur après l'étape de la détection des visages. Simplement, cet effet ne serait pas aussi rapide que nous l'avions suggéré au début de ce travail de thèse.

5.3.2 Un traitement basé sur les HFS malgré une suffisance statistique des BFS ?

Dans l'**Expérience 4**, nous nous sommes intéressés au rôle des fréquences spatiales dans la détection de visages émotionnels ou neutres. L'une des hypothèses que nous avions émises était que les visages en BFS seraient traités plus rapidement que les visages en HFS. Aussi, en BFS, nous nous attentions à un avantage des visages émotionnels apeurés en comparaison aux visages émotionnels joyeux. Les résultats que nous avons observés ne permettent pas de valider ces hypothèses. Au contraire, nous avons observé que les saccades vers les visages émotionnels ou neutres étaient effectuées plus rapidement en HFS qu'en BFS. Elles étaient aussi plus souvent dans la bonne direction, particulièrement lorsque le visage émotionnel était joyeux. Ainsi, en BFS, nous n'avons pas observé un avantage en faveur du traitement des visages apeurés, que ce soit sur les latences ou la proportion de saccades correctes.

Selon le modèle *coarse-to-fine* du traitement de l'information visuelle, l'information en BFS est traitée plus rapidement que l'information en HFS (Bar, 2003 ; Hegde, 2008 ; Kauffmann, Chauvin et al., 2015 ; Kauffmann et al., 2014 ; Musel et al., 2012 ; Schyns et Oliva, 1994). Dans l'**Expérience 4**, les performances en BFS étaient supérieures au hasard, ce qui signifie que l'information transmise par les BFS était suffisante pour effectuer la tâche. Or, nous avons observé des latences plus courtes avec des visages présentés en HFS. Une explication possible pour ce biais en faveur des HFS dans l'**Expérience 4** est que les BFS ne sont pas assez informatives pour désambiguïser rapidement le contenu émotionnel ou neutre des visages. Dans notre tâche, nous avons deux visages côte à côte, qui ont une structure globale et donc un contenu en BFS probablement très similaire. Pour différencier les deux en termes d'émotion, il est peut-être plus facile de s'appuyer sur des informations plus détaillées.

Cette idée serait en accord avec l'hypothèse d'une utilisation flexible des fréquences spatiales pour le décodage des expressions faciales. Différentes fréquences seraient extraites selon l'expression (Morrison et Schyns, 2001 ; Oliva et Schyns, 1997 ; Schyns et al., 2009 ; M. L. Smith et al., 2005) et la tâche (Schyns et Oliva, 1999 ; M. L. Smith et Merlusca, 2014). Par exemple, Schyns et Oliva (1999) ont constaté que lorsque les participants devaient catégoriser les expressions faciales comme étant en colère, joyeuses ou neutres, ils s'appuyaient davantage sur les informations BFS, alors qu'ils s'appuyaient sur les HFS lorsqu'ils devaient indiquer si le visage était émotionnel ou neutre. Ce résultat est cohérent avec nos données (qui montrent un avantage des HFS dans une tâche où les participants doivent discriminer un visage neutre et un visage émotionnel). Par conséquent, l'utilisation de l'information issue des différentes bandes de fréquences spatiales ne serait pas un processus fixe, mais dépendant de l'échelle des caractéristiques diagnostiques, en

considérant à la fois les contraintes de la tâche et la configuration spatiale du visage. Nous pouvons aussi supposer que les participants orientent au préalable leur attention vers des différences locales plutôt que globales. Plusieurs études montrent que l'attention peut être dirigée de manière flexible vers un niveau global et local en fonction des attentes des participants. Par exemple, des études utilisant des stimuli hiérarchiques comme amorces (par exemple des lettres globales formées de lettres locales) ont montré une amélioration du traitement des HFS après une orientation de l'attention vers la structure locale, et une amélioration du traitement des BFS après une orientation de l'attention vers la structure globale (Flevaris et al., 2011 ; Robertson et Ivry, 2000).

Dans l'ensemble, nos données n'appuient pas l'idée que les visages apeurés soient automatiquement mieux détectés par le système visuel, et ce sur la base d'un traitement des BFS qui serait indépendant des contraintes de la tâche. Une telle hypothèse était basée sur l'existence d'une voie sous-corticale pour la détection rapide des stimuli menaçants, incluant les visages apeurés, qui traiterait exclusivement les BFS (Méndez-Bértolo et al., 2016 ; Tamietto et de Gelder, 2010 ; Vuilleumier et al., 2003). Cependant, il est important de préciser que l'objectif de l'**Expérience 4** n'était pas d'apporter la preuve de l'existence de la voie sous-corticale. L'objectif était plutôt de tester si, au niveau comportemental, les visages apeurés pouvaient être détectés plus efficacement que d'autres visages sur la base du traitement des BFS. En considérant seulement les résultats de l'**Expérience 4**, nous ne pouvons pas exclure que les visages apeurés, contenant des BFS, activent l'amygdale plus tôt ou plus intensément. Nous avons néanmoins montré qu'un traitement privilégié de ces visages ne se reflète pas automatiquement sur la programmation des mouvements oculaires.

5.3.3 Attributs diagnostiques

Au cours de ces travaux de thèse, nous sommes aussi intéressés aux régions importantes pour la discrimination de visages neutres et émotionnels. Globalement, nos travaux ont clairement mis en évidence l'importance de la région de la bouche. D'abord, l'**analyse de la saillance** a montré que la saillance de la bouche, contrairement à celle des yeux, corrélait avec les performances obtenues dans les différentes conditions expérimentales de l'**Expérience 4**. Ce résultat suggère que les différences statistiques (en termes de différences de luminance, de fréquences spatiales et d'orientations) situées au niveau de la bouche contribuent à l'efficacité de la détection des visages émotionnels et neutres. Ensuite, les cartes de saillance du CNN obtenues lors de la la **simulation de l'Expérience 4** ont mis en évidence l'importance de cette région en particulier dans une tâche de discrimination de visages neutres et émotionnels. Ces résultats suggèrent que la région des yeux joue un rôle limité dans la détection de visages neutres et émotionnels, une idée qui peut paraître en opposition avec les études qui montrent que les yeux et la bouche sont deux régions très importantes (Dailey et al., 2002 ; Eisenbarth et Alpers, 2011 ; F. W. Smith et Schyns, 2009 ; Wegrzyn et al., 2017). Cependant, ces études concernent des tâches de catégorisation d'émotions précises. Elles ont montré que la bouche est particulièrement utile pour la catégorisation des visages joyeux, tandis que

les yeux sont particulièrement utiles pour la catégorisation des visages apeurés. Dans l'**Expérience 4**, la tâche est différente, et se rapproche plutôt d'une tâche de catégorisation de visages comme étant émotionnels ou neutres. Il n'est donc pas surprenant que les caractéristiques utilisées soient différentes de celles utilisées lors de tâches de catégorisation d'émotions précises. De plus, nous avons utilisé uniquement des visages joyeux et apeurés, ayant souvent la bouche ouverte. Les résultats pourraient être différents avec d'autres expressions, ou d'autres d'images.

5.3.4 Distribution de l'attention dans le visage pendant la programmation de saccades

De manière intéressante, l'étude des points d'arrivée des saccades dans chacune de nos expériences en choix saccadique (**Expériences 1, 2, 3, 4**) a révélé un schéma constant. En effet, ils arrivaient toujours plus bas lorsque la cible était un visage joyeux que lorsque la cible était un visage apeuré. Dans les (**Expériences 1 et 3**), ils arrivaient également plus bas lorsque la cible était un visage apeuré plutôt qu'un visage neutre. Bien que ces effets n'étaient pas au centre de ce travail de thèse, nous avions dès l'**Expérience 1** émis l'hypothèse selon laquelle les points d'arrivée arriveraient plus près de la bouche lorsque la cible est un visage joyeux que lorsqu'elle est un visage apeuré. Cette hypothèse venait de l'idée que la programmation des saccades se produit dans une carte de priorité organisée de manière rétinotopique (Belopolsky, 2015 ; Bisley et Mirpour, 2019 ; Fecteau et Munoz, 2006 ; Klink et al., 2014 ; Theeuwes, 2019). Chaque localisation de la scène se voit attribuer un poids en fonction, par exemple de sa saillance physique et de sa pertinence pour l'observateur. Les expressions faciales émotionnelles présentent différentes caractéristiques diagnostiques, principalement au niveau de la forme des yeux et de la bouche (Eisenbarth et Alpers, 2011 ; M. L. Smith et al., 2005 ; Wegrzyn et al., 2017). La région de la bouche est particulièrement importante pour le décodage des expressions joyeuses, et la région des yeux pour le décodage des expressions apeurées. Nous pouvons imaginer que cela se traduit par des poids différents dans la carte des priorités, plus forts pour les visages joyeux au niveau de la bouche, et plus forts pour les visages apeurés au niveau des yeux. En supposant que les points d'arrivée reflètent la distribution des poids dans la carte des priorités pendant la programmation des saccades, un poids plus important sur la région de la bouche se traduirait par une attraction des points d'arrivée vers la bouche.

Dans toutes nos expériences, les points d'arrivée se situaient entre la bouche et les yeux. Donc, parler de points d'arrivée plus bas signifie plus proches de la bouche. Les décalages observés en fonction des émotions étaient très faibles, et difficilement visibles en observant les cartes des points d'arrivée. Nous pouvons supposer que les points d'arrivée sont plus bas pour les visages joyeux du fait que leur bouche est ouverte, et donc particulièrement saillante. Pour les visages apeurés, certains ont la bouche fermée, mais la plupart ont la bouche ouverte, ce qui pourrait aussi la rendre particulièrement saillante en comparaison à la bouche des visages neutres, et expliquer pourquoi nous avons aussi observé des points d'arrivée plus bas pour les visages apeurés que neutres dans certaines expériences. Il est intéressant de noter que la tâche des participants module aussi les

points d'arrivée. Par exemple, les points d'arrivée étaient généralement situés autour des yeux dans l'**Expérience 1**, et autour du nez dans l'**Expérience 2**. Ces différences entre la première et la deuxième expérience peuvent être expliquées par les différentes tâches. Il est probable que lorsque les participants doivent traiter des expressions faciales, leur regard est naturellement plus orienté vers la bouche. Ces résultats sont en accord avec les résultats de la **simulation de l'Expérience 4** et de l'**analyse de saillance** qui souligne l'importance de cette région dans la détection de visages neutres et émotionnels. Dans l'**Expérience 3**, la position des points d'arrivée était intermédiaire.

En résumé, lorsque les participants programment une saccade vers un visage, nous pouvons supposer que l'attention est dirigée vers lui. Néanmoins, nous supposons que l'attention peut être dirigée à différents endroits du visage en fonction de la tâche et de l'expression, ce qui conduit à des différences au niveau des points d'arrivée. La modulation des points d'arrivée par les expressions peut être caractérisée d'automatique, puisque cet effet était observé dans toutes nos expériences. Néanmoins, elle ne traduirait pas l'intégration du contenu sémantique des expressions, mais seulement un traitement perceptif.

5.3.5 Implication de la voie sous-corticale dans la perception des expressions faciales

L'ensemble de ce travail de thèse a été guidé par l'idée de l'existence d'une voie sous-corticale, reliant le CS à l'amygdale en passant par le pulvinar, qui traiterait uniquement les BFS et qui permettrait une détection rapide des expressions faciales émotionnelles, en particulier apeurées. Nous supposions que cette information puisse être intégrée rapidement dans les mécanismes de programmation des saccades, et être à la base d'une capture automatique du regard par les visages émotionnels. Nos résultats comportementaux ne supportent pas l'idée qu'il existe un traitement privilégié des visages apeurés, fondé sur le traitement des BFS. Ils remettent aussi en question l'aspect automatique de la capture de l'attention et du regard par les visages émotionnels. Néanmoins, il est difficile de savoir sur la base de seuls résultats comportementaux quelles en sont les implications neurales. Par exemple, il est possible que la voie sous-corticale existe telle que nous l'avions définie, mais que les répercussions sur la programmation des saccades ne soient pas automatiques. L'une des études les plus connues qui apporte des données en faveur d'un traitement en double voie des expressions faciales est celle de Vuilleumier et al., (2003). Dans leur étude, les auteurs avaient étudié grâce à des enregistrements en IRMf l'activité neurale en réponse à des visages neutres ou apeurés présentés dans différentes bandes de fréquences spatiales. Ils ont observé des réponses plus fortes pour les visages apeurés que neutres au niveau de l'amygdale et du thalamus (un cluster qui selon les auteurs peut correspondre au pulvinar et au CS). Mais cette réponse était limitée aux images qui contenaient des BFS.

Dans l'**Expérience 5** nous avons tenté de reproduire une expérience similaire à celle de Vuilleumier et al. (2003), sans inclure les images non filtrées, et en ajoutant une condition de contraste égalisé. Néanmoins, même lorsque le contraste n'était pas égalisé (ce qui correspond à la configuration de l'étude de Vuilleumier et al., 2003), nous n'avons pas reproduit leurs résultats. Ainsi, que ce soit au niveau de l'amygdale, du pulvinar ou

du CS, nous n'avons pas observé de différences d'activations en fonction de l'expression faciale du visage, ni en BFS ni en HFS. Il est possible qu'un biais dans la création de nos stimuli nous empêche d'observer l'effet des émotions auquel nous nous serions attendus. En effet, nous avons essayé de nous rapprocher au mieux des conditions de l'étude de Vuilleumier et al. (2003). Nous avons utilisé des images issues, en partie, de la même base de visages. Plus précisément, dans leur étude deux bases différentes ont été utilisées, incluant la KDEF dont nous avons également fait usage. Ainsi, certaines images étaient identiques, mais pas toutes. Ensuite, nous avons recoupé et filtré les images en respectant les paramètres cités dans leur étude, par exemple les fréquences de coupures pour le filtrage. Néanmoins, visuellement nos stimuli semblent inclure légèrement plus de BFS. Sans certitude, nous pouvons seulement suggérer que ces différences viennent de la forme du filtre utilisé. Dans notre étude, nous avons utilisé un filtre gaussien pour obtenir nos images filtrées, avec lequel la partie exclue ne l'est jamais entièrement. Plus précisément, deux filtres gaussiens ont été créés, l'un laissant passer les HFS et l'autre les BFS, et ont été multipliés au spectre d'amplitude des images. Cette méthode a pour avantage d'éviter la création d'artefacts apparaissant avec un filtrage strict, dit *Heaviside*. Une alternative aux filtres gaussiens et *Heaviside* est le filtre *Butterworth*, qui produit des images plus fines que le filtrage gaussien, avec moins d'artefacts qu'un filtrage *Heaviside* (Dogra et Bhalla, 2014 ; Makandar et Halalli, 2015 ; Perfetto et al., 2020). Dans l'étude de Vuilleumier et al. (2003) la forme du filtre utilisé n'est pas détaillée.

Nous pouvons remarquer que nos images en BFS rendent les yeux des visages presque indiscernables. Or, plusieurs études ont mis en avant un lien fort entre l'information portée par la région des yeux et l'activité de l'amygdale (Adolphs et al., 2005 ; Ahs et al., 2014 ; Asghar et al., 2008 ; Kennedy et Adolphs, 2010 ; M. J. Kim et al., 2016 ; Meletti et al., 2012 ; Morris et al., 2002 ; Whalen et al., 2004). Par exemple, une étude de Ahs et al. (2014) suggère que l'activité de l'amygdale est corrélée aux déplacements des sourcils et à la proportion de blanc dans les yeux. En ce sens, d'autres études ont mis en avant une sensibilité de l'amygdale au blanc des yeux (M. J. Kim et al., 2016 ; Whalen et al., 2004), à la taille de la pupille (Demos et al., 2008) ou à l'orientation du regard (Adams Jr et al., 2003). Cette hypothèse d'un lien entre l'amygdale et le traitement des yeux est aussi supportée par des études chez un patient atteint d'une lésion bilatérale de l'amygdale qui témoignent d'une diminution du nombre de fixations sur la région des yeux (Adolphs et al., 2005 ; Kennedy et Adolphs, 2010). Ces données suggèrent que l'amygdale pourrait utiliser les yeux écarquillés comme une approximation grossière de la présence de visages apeurés, et offrent une explication perceptive (c'est-à-dire, basée sur l'intégration des caractéristiques physiques des stimuli) à un mécanisme impliquant le jugement de stimuli sociaux complexes (Whalen et al., 2013). Ainsi, si la réponse émotionnelle de l'amygdale est guidée par l'information située au niveau des yeux, et si nos stimuli en BFS les rendent indiscernables, cela peut expliquer l'absence de l'effet des émotions.

Au final, nos résultats remettent en question l'implication même de la voie sous-corticale dans la perception des émotions. Comme nous l'avons évoqué dans les paragraphes précédents, cela pourrait être dû à un biais méthodologique induit par le filtrage spatial. De

plus, l'utilisation de l'IRMf, dont la résolution temporelle est faible, peut nous empêcher de voir certains effets qui arrivent à un moment précis. Ces résultats se placent cependant en accord avec certaines études qui suggèrent que la voie sous-corticale serait finalement insensible aux expressions du visage. Notamment, certaines études ont utilisé la DCM pour évaluer les connexions entre les régions lors de la perception de visages exprimant différentes émotions à partir d'enregistrements de l'activité en MEG (Garvert et al., 2014 ; McFadyen et al., 2017). Dans ces études, les auteurs avaient comparé, entre autres, des modèles comportant seulement une voie corticale (reliant le CGL, V1 et l'amygdale), et des modèles comportant à la fois une voie corticale et une voie sous-corticale (reliant le pulvinar et l'amygdale). D'après leurs résultats, les modèles incluant la voie sous-corticale sont les plus probables. Néanmoins, cette voie ne serait influencée ni par les émotions du visage ni par les fréquences spatiales. De plus, certains modèles du traitement des visages suggèrent qu'une telle voie serait impliquée dans la détection des visages. Plus précisément, dans le cadre d'une revue de la littérature Johnson a proposé un modèle en double voie du traitement des visages : une voie sous-corticale pour la détection et une voie corticale pour l'identification (Johnson, 2005 ; Johnson et al., 2015). La voie sous-corticale serait rapide, opérerait sur la base d'une information en BFS et impliquerait le CS, le pulvinar et l'amygdale. Pour Johnson, cette voie expliquerait les résultats d'études chez les nourrissons, qui montrent que certaines informations sur les caractéristiques des visages sont disponibles dès la naissance. En effet, les nouveau-nés humains s'orientent préférentiellement vers des modèles schématiques simples ressemblant à des visages (Johnson et al., 1991). Ce modèle suggère que les activations plus prononcées pour les expressions apeurées pourraient venir du fait qu'ils constituent un modèle de visages pertinents pour la survie. Cependant, la fonction de la voie sous-corticale ne serait pas de détecter rapidement les émotions, mais de détecter rapidement les visages.

Nos résultats soulignent l'importance de régions corticales, telles que la FFA et l'OFA dans la discrimination des émotions. Ce qui est en accord avec les modèles du traitement cortical des expressions faciales (Liu et al., 2021). Le modèle multi-vagues proposé par Pessoa et Adolphs en 2010 accorde plus d'importance au traitement cortical que le modèle en double voie. Il suggère que le traitement de l'information émotionnelle ferait intervenir un réseau distribué, impliquant plusieurs voies différentes qui pourraient interagir entre elles. L'amygdale serait néanmoins toujours le noyau de l'évaluation de la pertinence biologique et pourrait moduler l'activité de ces différents réseaux. Il souligne aussi l'existence probable de très nombreuses voies à partir du pulvinar, vers des régions corticales ou sous-corticales. Dans une étude récente, Zhou et al. (2018) ont également mis en évidence le fait que les projections du pulvinar vers le cortex extrastrié exercent une forte influence sur les connexions entre le cortex extrastrié et l'amygdale. Ce qui suggère que le pulvinar pourrait moduler l'activité de l'amygdale via le cortex visuel plutôt que par une voie directe. Aussi, bien que plusieurs études ont montré une altération du traitement des stimuli émotionnels après une lésion de l'amygdale, d'autres études témoignent d'une réponse à des stimuli émotionnels même après une lésion de l'amygdale (Bach et al., 2015 ; Piech et al., 2011 ; Piech et al., 2010 ; Tsuchiya et al., 2009 ; Wang

et al., 2014). Certains auteurs proposent qu'en l'absence d'une amygdale fonctionnelle, la capture de l'attention par les émotions puisse reposer sur d'autres structures comme que les aires visuelles corticales ou le pulvinar (Bach et al., 2015). Finalement, toutes ces données remettent en question le rôle central de l'amygdale dans la médiation de la capture de l'attention par des stimuli émotionnels.

En résumé, nos données ne nous permettent pas de conclure précisément quant au réseau impliqué dans la capture du regard par les expressions faciales. Elles vont néanmoins à l'encontre de l'idée selon laquelle l'amygdale aurait un rôle central dans la perception des expressions faciales, et soulignent plutôt l'importance de régions corticales telles que la FFA et l'OFA. Dans nos expériences en choix saccadique, qui montrent une capture du regard plus importante par les visages émotionnels, les latences relativement longues des saccades (dès 150 ms, mais en moyenne autour de 250 ms si l'on se réfère aux résultats de l'**Expérience 2**), ainsi que l'importance des HFS, nous permettent de suggérer qu'un traitement cortical (impliquant donc la voie rétino-géniculo-striée) est plus probable. L'information serait ensuite transmise à l'OFA, puis à la FFA. Le traitement des visages dans la FFA pourrait débuter dès 130 ms (Hinojosa et al., 2015). En considérant un délai de 20 ms pour déclencher une saccade (Schiller et Kendall, 2004), il est ainsi possible que l'information soit transmise depuis le gyrus fusiforme même pour les saccades effectuées en 150 ms.

5.3.6 Bilan sur l'influence des expressions faciales dans la programmation des saccades

L'ensemble de ces travaux de thèse nous a permis de mieux comprendre l'influence des expressions faciales dans la programmation des saccades oculaires. Au vu des résultats que nous avons présentés et discutés dans les paragraphes précédents, nous pouvons caractériser l'influence des expressions faciales sur la programmation des saccades de la manière suivante. Lorsqu'un visage apparaît dans notre champ visuel, nous sommes capables de le détecter et de déclencher une saccade vers lui en 100-110 ms (comme en témoignent les mesures des latences minimales dans notre **Expérience 1** et dans d'autres études, par exemple Crouzet, 2010). Cette détection des visages se ferait à partir d'un traitement grossier de l'information, qui pourrait dépendre de l'analyse du spectre d'amplitude des visages (Honey et al., 2008) ou de l'analyse de parties isolées comme les yeux (Kauffmann et al., 2021). Nous pouvons suggérer qu'à ce niveau dans le décours temporel de la perception visuelle, les visages émotionnels ne sont pas privilégiés, c'est-dire qu'ils ne vont pas attirer plus vite ou plus souvent le regard que les visages neutres (**Expérience 1**). Néanmoins, les expressions faciales seraient perçues et impacteraient les points d'arrivée des saccades (**Expériences 1, 2, 3, 4**), ce qui n'implique pas nécessairement que le contenu émotionnel est décodé. Ainsi, dès 100 ms après l'apparition d'un visage, celui-ci aurait un poids fort dans la carte de priorité, en comparaison à d'autres objets. Nous pouvons suggérer que ce poids est réparti différemment en fonction de l'expression. Par exemple, si la bouche est ouverte le poids alloué à la région de la bouche serait plus fort que si elle est fermée, ce qui expliquerait les décalages observés au niveau des points d'arrivée.

Ensuite, dès 150 ms (comme en témoignent les mesures des latences minimales dans l'**Expérience 2**), nous sommes capables de détecter la présence d'un visage émotionnel. À partir de là, nous suggérons que les visages émotionnels peuvent bénéficier d'un traitement privilégié, en comparaison à des visages neutres. Plus précisément, ils vont avoir tendance à attirer le regard plus vite. Ce traitement privilégié se traduirait par un poids plus important pour les visages émotionnels que neutres dans les mécanismes de programmation des saccades. Nous supposons que cet effet peut être indépendamment de l'objectif des participants. Ainsi, nous avons observé un traitement privilégié des visages émotionnels même lorsque les participants devaient détecter le genre des visages (**Expérience 3**). Nous avions émis l'hypothèse que le traitement privilégié des visages émotionnels puisse être mis en place rapidement par l'intermédiaire d'une voie sous-corticale, qui relie le CS à l'amygdale en passant par le pulvinar. Cette voie est censée être particulièrement activée par des visages apeurés, et traiterait uniquement les BFS. Nos résultats ne permettent pas d'appuyer cette hypothèse pour plusieurs raisons. D'abord, la détection des expressions faciales au niveau comportemental ne favorise pas les visages apeurés, en BFS (**Expérience 4**). Ensuite, les temps de réaction observés dans les expériences comportementales qui montrent un traitement privilégié des visages émotionnels sont assez longs pour faire intervenir la voie corticale classique. Ainsi, nous pouvons envisager l'implication de la voie rétino-géniculo-striée, par laquelle l'information serait transmise à l'OFA, puis à la FFA. Le traitement des visages et des expressions dans le gyrus fusiforme est connu pour atteindre un pic aux alentours de 170 ms, mais pourrait débuter dès 130 ms (Hinojosa et al., 2015). En considérant un délai de 20 ms pour déclencher une saccade (Schiller et Kendall, 2004), il est possible que l'information soit transmise depuis le gyrus fusiforme. Pour finir, nos résultats en IRMf ont souligné des activations plus fortes dans ces régions pour les visages apeurés que neutres, ce qui n'était pas le cas dans l'amygdale (analyse en ROI de l'**Expérience 5** ; notons néanmoins que, bien que cet effet ne soit pas assez fort pour se refléter dans l'analyse en ROI, les cartes des activations ont mis en avant l'implication d'une petite partie de l'amygdale dans le traitement des émotions). Un résumé des principaux résultats obtenus en lien avec les mécanismes de programmation des saccades est présenté sur la Figure 5.1. Les temps qui sont donnés ici, 100 ms pour la détection des visages et 150 ms pour la détection des visages émotionnels, sont des temps minimaux. En moyenne, ces étapes arriveraient plus tardivement (176 ms pour la détection des visages et 249 ms pour la détection des visages émotionnels, d'après les **Expériences 1 et 2**).

5.4 Apports méthodologiques

Dans cette section, nous allons revenir sur les apports méthodologiques de ce travail de thèse. Nous discuterons d'abord de nos résultats concernant l'impact de l'égalisation ou non du contraste lors de l'utilisation d'images filtrées. Ensuite, nous reviendrons sur l'intérêt des réseaux de neurones artificiels dans l'étude du comportement humain mis en avant dans nos travaux.

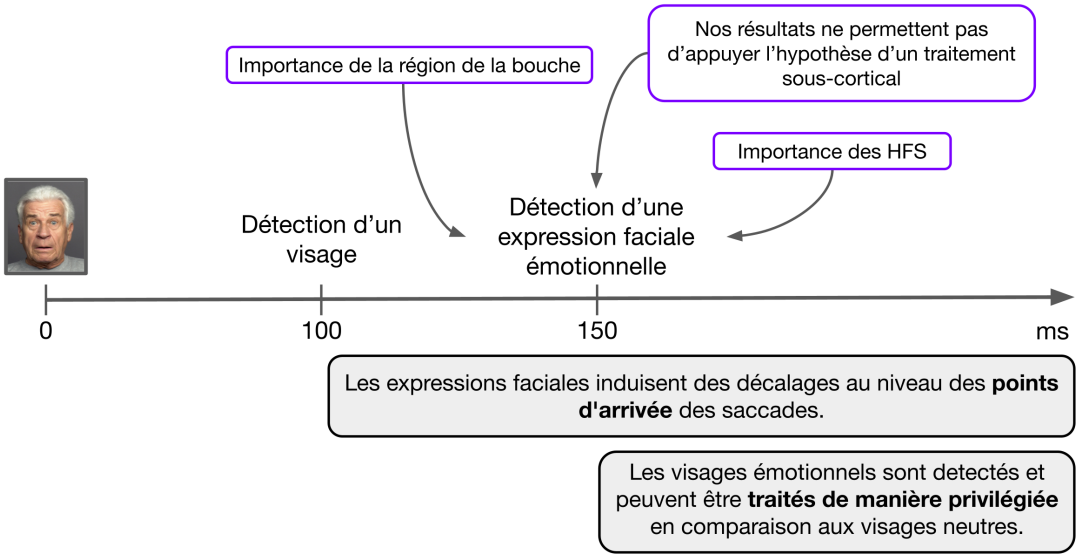

Figure 5.1 – Bilan des résultats obtenus sur l'influence des expressions faciales dans la programmation des saccades.

5.4.1 Effet de l'égalisation du contraste d'images filtrées

Au cours de nos travaux, nous avons eu l'occasion de tester l'effet de l'égalisation du contraste entre les fréquences spatiales sur la détection et la programmation de saccades vers des visages neutres et émotionnels (**Expérience 4**), ainsi que sur l'activité cérébrale induite par la perception de visages neutres ou apeurés (**Expérience 5**). Pour rappel, après un filtrage des fréquences spatiales, les BFS possèdent naturellement un contraste plus élevé que les HFS. Cette propriété découle du fait que le contraste décroît avec l'augmentation des fréquences spatiales. Certains auteurs choisissent de conserver ces différences de contrastes. Mais, en conservant ces propriétés naturelles, il est difficile de dissocier un effet dû à des différences de fréquences spatiales d'un effet dû à des différences de contraste (par exemple un stimulus plus contrasté aura tendance à être détecté plus facilement). Ainsi, afin d'isoler l'effet des fréquences spatiales de celui du contraste, d'autres auteurs choisissent d'égaliser le contraste après le filtrage spatial.

Dans l'**Expérience 4**, l'interaction attendue entre le contraste et les fréquences spatiales n'a pas été observée sur les proportions de saccades correctes, mais elle a été observée de manière marginalement significative sur les latences des saccades. Plus précisément, les saccades étaient effectuées plus rapidement en HFS qu'en BFS. Cet effet était seulement significatif lorsque le contraste était égalisé. Bien que cette interaction n'ait pas atteint le seuil de significativité, nous pouvons souligner que la direction de l'effet est en accord avec les résultats de la littérature (Kauffmann, Chauvin et al., 2015 ; Kauffmann, Ramanoël et al., 2015 ; Perfetto et al., 2020 ; Vlamings et al., 2009). Par exemple, dans une étude réalisée par Vlamings et al. (2009) les participants devaient décider si le visage présenté était apeuré ou neutre. Les visages étaient présentés en BFS ou en HFS et avec

ou sans égalisation du contraste. Les résultats ont montré que les visages en BFS étaient catégorisés plus rapidement que les visages en HSF, pour les stimuli égalisés et non égalisés. Cependant, cette différence était plus forte lorsque le contraste n'était pas égalisé. Nous pouvons suggérer que le fait que notre interaction sur les latences ne soit que marginalement significative est dû à un manque de puissance. Concernant la proportion de saccades correctes, cette mesure n'était pas sensible aux différentes conditions de contraste. Nous pouvons suggérer que le contraste global n'est pas important pour discriminer précisément les visages émotionnels et neutres. Il est aussi possible que, du fait que nos conditions de fréquences spatiales ont été présentées en blocs, les participants se soient habitués au contraste du bloc et l'ont adopté comme une ligne de base. Dans l'ensemble, les effets du contraste étaient moins importants que ce à quoi nous nous attendions.

Dans l'**Expérience 5**, l'effet du contraste était également moins important que prévu. Plus précisément, bien que l'égalisation du contraste était associée à un agrandissement des clusters occipito-temporaux qui répondaient spécifiquement aux HFS, les analyses en ROI n'ont pas montré d'interaction entre les fréquences spatiales et le contraste. La plupart des régions étudiées (c'est-à-dire l'amygdale droite et gauche, latérale ou médiale, ainsi que le pulvinar gauche, la FFA et l'OFA) ont montré des réponses plus fortes face à des visages en HFS que des visages en BFS. Mais, cet effet n'était pas plus fort avec l'égalisation du contraste.

Dans l'ensemble, nos résultats ont mis en avant un impact du contraste dans le traitement de visages filtrés aussi bien au niveau comportemental que neural. Cependant, les effets obtenus étaient minimes, puisqu'ils consistaient uniquement en une interaction marginale sur les latences des saccades dans l'**Expérience 4**, et une augmentation de la taille de certains clusters dans l'**Expérience 5**. Pour la plupart des variables étudiées, nous n'avons observé aucun effet de l'égalisation du contraste.

5.4.2 Utilisation des réseaux de neurones artificiels dans le cadre de l'étude du comportement humain

Par l'intermédiaire de la **simulation de l'Expérience 2**, ainsi que de la **simulation de l'Expérience 4**, nous avons tenté de mettre en avant l'intérêt des réseaux de neurones artificiels dans l'étude du comportement humain. Les deux simulations nous ont permis d'aborder différentes perspectives. Dans la **simulation de l'Expérience 2**, l'intérêt était de comparer les performances des participants avec celle d'un système qui n'est pas soumis à l'interprétation du contenu émotionnel des images. Ainsi, nous ne prétendons pas que ce modèle représente précisément les étapes impliquées dans le traitement des expressions faciales. Le réseau était plutôt considéré comme un outil qui permet de quantifier les différences statistiques entre un visage joyeux et un visage neutre, ou entre un visage apeuré et un visage neutre. Une utilisation similaire des réseaux de neurones a déjà été présentée dans certaines études (Dailey et al., 2002 ; Mermillod et al., 2009). Par exemple, Dailey et al. (2002) ont mis en avant une similarité entre la reconnaissance d'émotions chez les humains et les réseaux de neurones. Leurs résultats ont montré que, aussi bien pour les humains que pour le réseau, la peur était l'émotion la plus difficile à reconnaître,

tandis que la joie était l'émotion la plus facile à reconnaître. Globalement, ces résultats comme les nôtres suggèrent que la reconnaissance des émotions peut reposer uniquement sur les propriétés physiques des images. Dans l'**Expérience 2**, nous n'étions pas dans une configuration classique de catégorisation d'expressions faciales, puisque les participants devaient discriminer deux images. Nous avons adapté notre modèle à ce contexte, et montré qu'aussi bien pour les humains que pour le réseau, la discrimination entre un visage émotionnel et un visage neutre était plus facile avec un visage émotionnel joyeux (en comparaison à un visage émotionnel apeuré). En résumé, ces résultats viennent soutenir l'idée que les réseaux de neurones artificiels peuvent être utilisés pour tester la possibilité que des résultats comportementaux soient expliqués par des facteurs statistiques.

Dans la **simulation de l'Expérience 4**, l'intérêt de l'utilisation d'un réseau de neurones était de proposer une méthode de visualisation des régions utiles à une tâche. La méthode utilisée dans le cadre de cette simulation (Simonyan et al., 2014) n'est pas nouvelle, dans le sens où elle est déjà utilisée dans le domaine de l'intelligence artificielle. Ici, nous l'avons adapté à notre problématique et nous suggérons qu'elle peut aussi avoir sa place dans le champ de l'étude du comportement humain. L'objectif de cette étude était de mettre en évidence les régions qui sont diagnostiques pour la discrimination d'un visage émotionnel et neutre, et de tester si cette information permet de prédire les résultats de l'**Expérience 4**. Dans la littérature, plusieurs méthodes permettent d'identifier des caractéristiques utiles pour des tâches spécifiques, par exemple des tâches de reconnaissance des expressions faciales. Une méthode directe passe par l'enregistrement des mouvements oculaires pendant que les participants catégorisent l'expression des visages. Dans notre expérience comportementale, les participants prenaient leur décision avec leur regard au centre de l'écran, puis déclenchaient une saccade vers la cible. Par conséquent, nous ne pouvions pas identifier avec l'enregistrement des mouvements oculaires les régions utiles du fait que les participants ne pouvaient pas explorer les images. Lorsque l'information issue des mouvements oculaires n'est pas disponible, il est intéressant d'avoir un outil qui met en avant les régions les plus utiles d'un point de vue statistique. Notre modèle a mis en évidence l'importance de la bouche dans la discrimination entre un visage neutre et un visage émotionnel. Bien que nous ne pouvons pas être certains que les participants ont utilisé les mêmes informations que le réseau, cela semble très probable. En effet, nous avons montré que la saillance de la bouche, et non celle des yeux, permet de prédire les performances des participants, ce qui suggère qu'ils utilisent cette information en particulier.

5.5 Limites et perspectives

Ce travail de thèse nous a permis d'éclaircir certains points concernant l'influence des expressions faciales sur la programmation des saccades. Néanmoins, il comporte également un certain nombre de limites, déjà partiellement abordées dans la discussion des différents chapitres expérimentaux. Nous reviendrons dans cette section sur les différentes limites de nos travaux, ainsi que sur des perspectives envisagées.

5.5.1 Généralisation à d'autres expressions émotionnelles

L'une des premières limites à laquelle nous pouvons penser vient du fait que nous nous sommes limités à l'étude de deux expressions faciales émotionnelles : la joie et la peur. Bien que le choix de ces deux expressions était motivé par les résultats de la littérature, il rend incertaine la généralisation de nos résultats à d'autres expressions émotionnelles. Comme nous l'avons vu dans le Chapitre 1, six émotions de base sont classiquement distinguées : la joie, la peur, la colère, la surprise, la tristesse et le dégoût. Elles sont chacune caractérisées par des configurations spécifiques, qui permettent de les différencier les unes des autres. La joie est connue pour être une expression facile à reconnaître, tandis que la peur est connue pour être difficile à reconnaître. Dans nos expériences en choix saccadique il est possible que les participants développent un modèle des visages émotionnels basé sur un mélange d'expressions joyeuses et apeurées. Ils pourraient décider d'adopter la stratégie de se concentrer sur la bouche parce qu'ils considèrent qu'il s'agit de la région la plus informative dans ce contexte, ce qui pourrait favoriser la détection des visages émotionnels joyeux. Ce modèle pourrait être différent avec d'autres expressions. Nous pouvons supposer que, les visages émotionnels auraient toujours tendance à être détectés et à attirer le regard plus rapidement, puisque cet effet a été observé même avec l'une des expressions les moins reconnaissables. En fonction des émotions utilisées, les performances pourraient être plus ou moins bonnes. La **simulation de l'Expérience 4** et l'**analyse de saillance** ont mis en avant l'importance de la bouche dans la tâche. Il est possible que, même avec une plus grande variété d'émotions, cet indice reste aussi pertinent. Cependant, nous pouvons supposer que si nous n'avions utilisé que des visages émotionnels pour lesquels la bouche n'est pas nécessairement ouverte, les résultats auraient été différents. Par exemple, avec des visages en colère, apeurés, ou tristes, ou avec des visages joyeux dont la bouche et fermée, les indices diagnostiques pourraient se situer également ou exclusivement au niveau des yeux. L'utilisation des cartes de saillance issues d'un CNN avec de tels stimuli permettrait de mettre en avant les régions utiles avec différents types de stimuli.

5.5.2 Temps de présentation des stimuli

Une autre limite concerne la durée de présentation de nos stimuli. En effet, certaines études suggèrent que la durée de présentation des images influence l'utilisation des différentes bandes de fréquences spatiales (Goffaux et al., 2011 ; Schyns et Oliva, 1994). Plus précisément, ces études suggèrent qu'il y aurait un biais de traitement en faveur des HFS pour les stimuli présentés relativement longtemps. Et, au contraire un biais de traitement en faveur des BFS pour les stimuli présentés brièvement. Par exemple, dans l'étude de Schyns et Oliva (1994) avec des images hybrides, un temps de présentation très court (30 ms) suscitait une catégorisation basée sur le contenu en BFS, alors qu'un temps de présentation plus long (150 ms) suscitait une catégorisation basée sur le contenu en HFS. Au niveau neural, Goffaux et al. (2011) ont montré que la présentation de visages en BFS et non filtrés pendant 75 ms évoquait une activité plus forte dans la FFA droite que celle de visages en HFS. Pour un temps de présentation plus long (150 ms),

le schéma était inversé, les visages présentés en HFS et non filtrés engageaient la FFA droite plus fortement que les visages présentés en BFS. Dans nos expériences, les temps de présentation peuvent être considérés comme longs (400 ms dans l'**Expérience 1**, 800 ms dans les **Expériences 2, 3, 4**, et 200 ms dans l'**Expérience 5**). Ainsi, il n'y avait pas de contrainte temporelle forte, ce qui pourrait induire un biais en faveur du traitement des HFS. Des études complémentaires peuvent être envisagées pour évaluer à quel point nos résultats sont dépendants du temps de présentation des images. Notons néanmoins que, dans une tâche de choix saccadique opposant des visages et des véhicules, un biais en faveur des BFS a été mis en avant avec des temps de présentation assez longs (400 ms ; ce biais était néanmoins seulement observé sur la mesure des latences minimales ; Guyader et al., 2017). Cela pourrait suggérer que, si le contenu porté par les BFS est suffisamment diagnostique, il pourrait être favorisé indépendamment du temps de présentation.

5.5.3 Paramètres du filtrage spatial

Ensuite, comme nous l'avons abordé plus tôt dans ce manuscrit, certains de nos résultats peuvent être dépendants des paramètres utilisés lors du filtrage spatial. Dans la littérature sur le traitement des fréquences spatiales dans un visage, les résultats sont parfois hétérogènes, ce qui peut au moins en partie s'expliquer par la variété des méthodologies utilisées (Jeantet et al., 2018 ; Perfetto et al., 2020). Dans l'**Expérience 4**, les fréquences de coupures correspondant aux BFS et aux HFS étaient de 11 cpi (1 cpd) et de 66 cpi (6 cpd), respectivement. Ces paramètres ont été choisis afin de se rapprocher des conditions de l'étude en choix saccadique avec des visages filtrés menée par Guyader et al. (2017). Dans l'**Expérience 5**, les fréquences de coupure étaient fixées à 6 cpi (1 cpd) pour les stimuli en BFS et à 24 cpi (4 cpd) pour les stimuli en HFS. Des paramètres qui avaient cette fois été choisis afin de se rapprocher des conditions de l'étude menée par Vuilleumier et al. (2003). Ainsi, nous pouvons déjà remarquer que nos stimuli en HFS dans l'**Expérience 4** étaient composés des fréquences plus hautes que dans l'**Expérience 5**. Pour les deux études, nous avons utilisé un filtre gaussien, qui peut laisser passer certaines fréquences non désirées et rendre nos stimuli HFS moins fins que d'autres types de filtres (Perfetto et al., 2020). Le fait que nous n'avons pas observé d'effet des émotions en BFS dans l'**Expérience 5** pourrait être dû au fait que nos stimuli sont trop flous, et rendent la région des yeux difficile à distinguer. Cette hypothèse pourra être testée dans de futures simulations, en utilisant des cartes de saillance issues d'un CNN par exemple.

En d'autres termes, une perspective envisagée serait d'utiliser un réseau de neurones artificiel tel que ceux que nous avons présentés précédemment (**simulation des Expériences 2 et 4**) afin d'évaluer à quel point les visages apeurés et neutres de l'**Expérience 5** sont distinguables, en HFS et en BFS. L'hypothèse sous-jacente est que la catégorisation des visages comme neutres ou apeurés serait plus facile en HFS. De plus, en utilisant seulement les HFS nous pouvons supposer que les yeux seraient utiles. Mais, en utilisant les images en BFS, il est possible qu'ils ne puissent plus être utilisés du fait qu'ils sont visuellement difficiles à distinguer. À plus long terme, il serait intéressant d'évaluer l'effet des émotions sur l'amygdale dans différentes bandes de fréquences, en utilisant des intervalles très fins. Cela permettrait d'identifier à partir de quelles bandes de fréquences précises l'amygdale permet de distinguer les émotions, et jusqu'à quelles bandes de fréquences.

5.5.4 Distinction entre égalisation physique et égalisation perceptive

Une autre limite qui n'a pas encore été abordée dans ce travail de thèse, et qui pourrait expliquer certains résultats obtenus, est liée à la distinction entre le stimulus distal et le stimulus proximal. Tout stimulus visuel présent physiquement dans notre environnement (le stimulus distal, par exemple une pièce), n'est pas traité en tant que tel, mais il est encodé par notre organisme de manière subjective en fonction de nos contraintes sensorielles. Le résultat de cet encodage est le stimulus proximal (par exemple, la lumière réfléchie par la pièce sur la rétine). Ainsi, une pièce sera toujours physiquement la même, mais pourra être perçue différemment en fonction de l'angle de vue, de la lumière ou de l'endroit où elle se trouve.

Dans ces travaux de thèse, lorsque nous manipulons le contraste de nos images, nous manipulons le stimulus distal ; le contraste réel est ainsi égalisé. Cependant, il n'est pas certain que le contraste perçu, celui du stimulus proximal, soit lui aussi égalisé. Au contraire, plusieurs études suggèrent que la sensibilité au contraste n'est pas la même pour toutes les bandes de fréquences. Plus précisément, la plupart de ces études génèrent une fonction de sensibilité au contraste, qui décrit la façon dont la sensibilité au contraste varie avec la fréquence spatiale. La plupart du temps, cette fonction est générée en utilisant des réseaux sinusoïdaux comme stimuli, et en estimant pour une large gamme de fréquences spatiales le seuil à partir duquel le contraste n'est plus visible par différents participants (par exemple, Campbell et Robson, 1968 ; De Valois et al., 1974 ; Lesmes et al., 2010, pour une revue de la littérature voir Pelli et Bex, 2013). Bien que la forme de la courbe obtenue puisse varier (par exemple, en fonction des participants ou de la luminance), elle semble généralement atteindre un pic aux alentours de 4 cpd. C'est-à-dire que pour cette fréquence spatiale, les différences de contrastes sont plus faciales à percevoir. Ensuite, plus on va vers les hautes ou les basses fréquences, plus la sensibilité diminue (c'est–à-dire que l'on va avoir besoin d'un contraste physique plus fort pour détecter un changement de luminance). Il est possible que nos images HFS, même avant toute égalisation, aient un contraste perçu identique ou plus fort que nos images en BFS, bien que le contraste réel soit plus fort pour les BFS. Le fait que les HFS puissent être mieux perçues malgré un plus faible contraste peut être dû à certains mouvements oculaires fixationels, qui augmenteraient la discriminabilité des stimuli en HFS. Ainsi, Rucci and Poletti (2015) ont montré que, dans une tâche de discrimination d'orientation de réseaux sinusoïdaux en HFS et en BFS, une stabilisation rétinienne (qui fait en sorte que l'image rétinienne soit stable en bougeant l'image en même temps que les yeux) fait baisser les performances en HFS, mais pas en BFS.

Si nous reprenons les valeurs utilisées dans l'**Expérience 4**, et que nous les comparons avec une courbe de sensibilité au contraste (ici issue du modèle de Barten, 2003), il semble que le contraste des HFS soit plus visible que celui des BFS. Ce constat est présent avant même que le contraste soit égalisé, et s'accentue donc après l'égalisation (Figure 5.2). La différence est encore plus forte en reprenant les valeurs de l'**Expérience 5**. En résumé, bien que la méthode d'égalisation du contraste que nous avons utilisé, qui

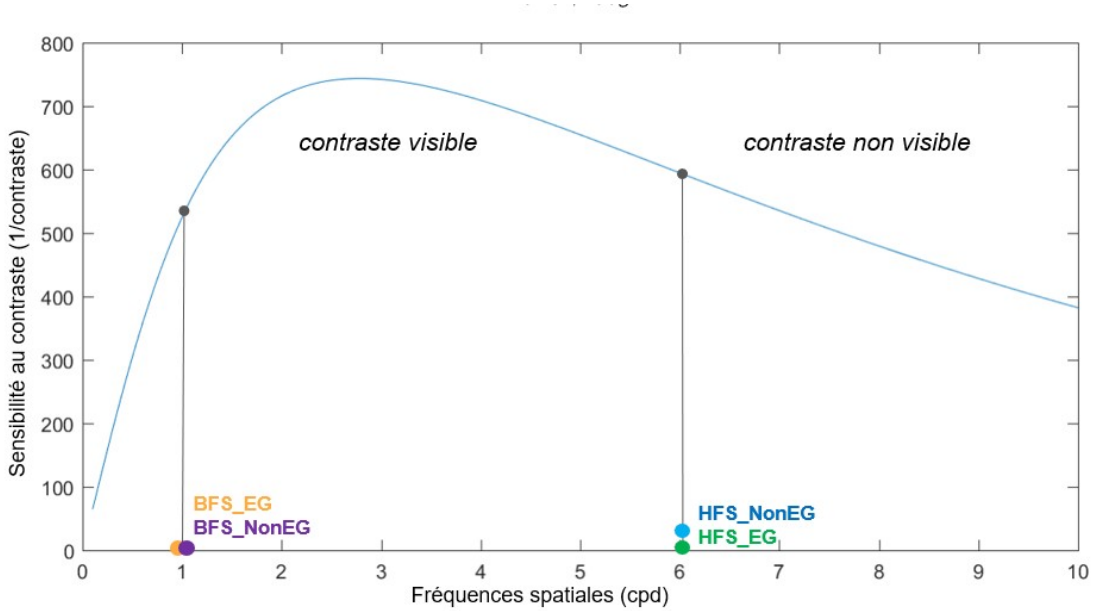

Figure 5.2 – Fonction de sensibilité au contraste issue du modèle de Barten (2003). La sensibilité au contraste calculée pour nos images en BFS et HFS, avant (NonEG) ou après (EG) l'égalisation dans l'Expérience 4 est affichée. Nous pouvons voir que, pour toutes nos images le contraste est visible. D'après ce modèle, les images les plus visibles (la visibilité est quantifiée en termes de distance au seuil du visible, voir les lignes grises sur la figure) sont dans l'ordre : les HFS-EG (589), les HFS-NonEG (562), les BFS-NEG (521) et les BFS-EG (519). Notons néanmoins que les différences sont faibles.

induit une égalisation physique, soit motivée par la littérature (car les études égalisant le contraste utilisent cette méthode), elle n'induit probablement pas une égalisation dans le domaine proximal. De plus, l'idée que les HFS seraient désavantagées par leur faible contraste physique est discutable, car elles pourraient être perçues comme autant ou même plus contrastées que les BFS en fonction des fréquences de coupures choisies.

5.5.5 Vision périphérique

Dans nos expériences comportementales (**Expériences 1, 2, 3, 4**), les visages étaient présentés en vison parafovéale. Le centre de chaque image était situé à 8° d'angle visuel du centre de l'écran et les images mesuraient 11° d'angle visuel. Il restait 1° d'angle visuel entre le centre et le début de chaque image. En condition naturelle, les visages apeurés sont plus susceptibles d'apparaître dans le champ visuel périphérique. Ainsi, nous pouvons nous demander à quel point nos résultats peuvent s'étendre à des stimuli présentés en vision périphérique. Plusieurs études ont montré que les expressions faciales peuvent être détectées en vision périphérique jusqu'à 30° ou 40° d'excentricité (Bayle et al., 2011 ; F. W. Smith et Rossit, 2018). Ces effets semblent particulièrement s'appliquer à la joie, la surprise et la peur (F. W. Smith et Rossit, 2018). Néanmoins, les performances baissent avec l'augmentation de l'excentricité. Dans nos expériences en choix saccadique, dans lesquelles le taux d'erreur et assez élevé, la tâche serait être encore plus difficile avec des stimuli présentés en périphérie. Mais, nous pouvons supposer que la

meilleure détection des visages émotionnels subsisterait jusqu'a de tels degrés excentricités. Néanmoins, l'étude de l'utilisation des fréquences spatiales pourrait être biaisée en faveur de BFS, les HFS étant difficiles à percevoir en vision périphérique. Au niveau cérébral, certaines études montrent que les expressions faciales évoquent des réponses différentes jusqu'à 30° d'excentricité (Rigoulot et al., 2011 ; Rigoulot et al., 2012). Dans une étude en MEG, Bayle et al. (2009) se sont intéressés aux réponses évoquées par des visages apeurés ou neutres, non consciemment perçus et présentés en vision centrale ou périphérique. Les résultats ont montré qu'une partie du gyrus temporal droit comprenant l'amygdale répondait aux visages apeurés entre 80 et 130 ms en vision périphérique, un effet qui n'était pas observé en vision centrale. Dans l'**Expérience 5**, nous n'avons pas observé d'effet des émotions au niveau de l'amygdale. Comme nous l'avons suggéré précédemment, cette absence d'effet pourrait s'expliquer par le fait que les yeux sont difficiles à distinguer en BFS. Il est néanmoins possible que ces effets soient plus susceptibles d'être observés en vision périphérique qu'en vision centrale.

5.5.6 Stimuli dynamiques

En conditions naturelles, les visages émotionnels apparaissent de manière dynamique, un aspect qui n'est pas considéré dans nos travaux. Bien que plus faciles à contrôler que les stimuli dynamiques, les stimuli statiques sont moins écologiques. Nous pouvons nous demander à quel point nos résultats peuvent s'étendre à des stimuli dynamiques. Certaines études ont suggéré que l'utilisation de stimuli dynamiques améliore le traitement des visages apeurés, ou des stimuli menaçants en général. Par exemple, une étude de Sato et al. (2004) a montré que l'activité évoquée par des visages apeurés dans l'amygdale gauche est plus forte lorsqu'ils sont dynamiques que statiques, un effet qui n'est pas observé avec des visages joyeux. Des régions occipito-temporales incluant le gyrus fusiforme montraient un pattern similaire (c'est-à-dire des activations accrues avec des stimuli dynamiques), que ce soit pour la joie ou la peur. Dans le même sens, une étude EEG a montré que les stimuli dynamiques menaçants (par exemple les araignées) provoquaient une réponse plus élevée que les stimuli statiques menaçants ou les stimuli dynamiques non menaçants au niveau de la composante P1 (Carretié et al., 2009). Ces résultats suggèrent que le mouvement apporte une saillance supplémentaire aux stimuli menaçants et facilite leur détection. Cependant, il n'existe encore que quelques études sur ce sujet et les résultats sont parfois contradictoires. Ainsi, certains résultats témoignent d'une activité accrue au niveau de l'amygdale pour les visages dynamiques indépendamment de l'émotion (Trautmann et al., 2009). D'autres études ont observé des activations similaires face à des visages dynamiques ou statiques au niveau de l'OFA, de la FFA ou de l'amygdale (Bernstein et al., 2018 ; van der Gaag et al., 2007).

Au niveau comportemental, des études utilisant des stimuli statiques similaires à ceux que nous avons utilisés dans nos expériences (c'est-à-dire des émotions de base avec une représentation de l'émotion à son pic d'intensité) n'ont pas observé un avantage des stimuli dynamiques dans le cadre de tâches de reconnaissance des émotions (Fiorentini et Viviani,

2011 ; Gold et al., 2013 ; Krumhuber et al., 2021). D'autres études ont observé un tel effet, mais en particulier dans des conditions où l'information statique n'était pas optimale (Dobs et al., 2018 ; Horstmann et Ansorge, 2009). Dans nos études, nous pouvons supposer que l'utilisation de stimuli statiques rend la tâche plus facile, car ils présentent directement l'information la plus pertinente (c'est-à-dire l'émotion à une intensité maximale).

5.5.7 Vers une caractérisation plus précise des conditions nécessaires à la capture du regard par les visages émotionnels

Nos expériences comportementales ont montré que la capture de l'attention et du regard par les expressions faciales émotionnelles n'est pas nécessairement dépendante de leur pertinence pour la tâche. Néanmoins, nous ne pouvons pas suggérer que cet effet s'applique pour toutes les tâches après l'étape de la détection des visages. Par exemple, dans des tâches dans lesquelles les visages ne sont pas pertinents ou d'autres tâches dans lesquelles les expressions faciales ne sont pas pertinentes. En effet, dans toutes nos tâches, les visages étaient pertinents. Ainsi, nous ne pouvons pas être sûrs que nos résultats s'étendent à des tâches dans lesquelles les visages sont non pertinents. En utilisant des réponses manuelles, certaines études n'ont pas observé de traitement privilégié des visages émotionnels, même sur des comportements tardifs, et soulignent l'aspect conditionnel de la capture de l'attention par les visages émotionnels. Par exemple, des tâches dans lesquelles un visage émotionnel est brièvement présenté en amorce, suivi d'un stimulus à catégoriser (Koster et al., 2007 ; Puls et Rothermund, 2018 ; Victeur et al., 2019). Dans une série de quatre expériences, Tannert et Rothermund (2020) ont présenté simultanément un visage cible et un ou plusieurs visages distracteurs à des participants. Ils ont observé un effet des émotions seulement pour une expérience, dans laquelle deux visages étaient présentés et les participants devaient indiquer si le visage cible (masculin ou féminin) était émotionnel ou neutre. Lorsque les participants devaient indiquer si le visage cible était vieux ou jeune, ou indiquer si un visage présenté en vision centrale et entouré de distracteurs était émotionnel ou neutre, cet effet n'était plus observé. Les auteurs ont suggéré que les visages émotionnels distracteurs vont capturer l'attention seulement lorsque le traitement des expressions faciales et le traitement de tous les stimuli sont obligatoires. Dans l'**Expérience 3**, nous avons observé un effet des émotions dans une tâche dans laquelle leur traitement n'était pas obligatoire. Néanmoins, cela peut être restreint au traitement du genre, ou à l'utilisation de réponses saccadiques qui peuvent être plus adaptées pour étudier des effets émotionnels (Bannerman, Milders et Sahraie, 2009).

Nous pouvons envisager de nouvelles expériences avec des réponses saccadiques afin de mieux comprendre les conditions nécessaires à l'observation d'une capture du regard par les visages émotionnels. Nous avons suggéré que les latences des réponses peuvent expliquer l'absence d'une capture du regard plus importante pour les visages émotionnels que neutres. Nous pouvons envisager de tester cette hypothèse en reprenant la procédure de l'**Expérience 1**, mais en obligeant les participants à attendre avant de répondre. Aussi, nous pouvons envisager des tâches dans lesquelles les visages sont non pertinents. Par exemple, une tâche de choix saccadique opposant des animaux et des véhicules, avec des

visages émotionnels ou neutres non pertinents pouvant apparaître à différents endroits de l'écran. Afin de nous rapprocher des conditions de l'expérience de Tannert et Rothermund (2020), nous pourrions aussi tester l'impact des expressions faciales dans une tâche de détection d'âge (aller vers le visage le plus vieux ou le plus jeune), dans laquelle il y a moins d'attentes stéréotypées concernant les expressions faciales.

5.5.8 Vers une comparaison plus fine des attributs diagnostiques pour les humains et pour le CNN

Pour finir, nous pouvons envisager plusieurs perspectives de travail en lien avec le modèle computationnel présenté dans la **simulation de l'Expérience 4**. À court terme, il serait intéressant d'utiliser ce modèle pour générer des cartes de saillance associées à une tâche de discrimination de genre, que ce soit avec des paires de visages (et ainsi simuler l'**Expérience 3**) ou avec des images seules (et ainsi simuler l'**Expérience 5**). Cela permettrait de visualiser les régions utiles dans nos différentes expériences, et les comparer avec celles que nous avons déjà obtenues en simulant l'**Expérience 4**. Comme nous l'avons évoqué précédemment, nous pouvons aussi envisager d'utiliser ce modèle pour évaluer à quel point les visages apeurés et neutres de l'**Expérience 5** sont distinguables, en HFS et en BFS. Cela permettrait aussi de quantifier l'importance des yeux, en HFS et en BFS.

Sinon, l'une des limites du modèle est qu'il ne simulait pas exactement la tâche des participants, puisque nous n'avons pas pris en compte les contraintes d'une réponse saccadique. Avec une architecture très simple, le CNN a atteint de hautes performances dans la discrimination des visages émotionnels et neutres. La performance globale du réseau dans les différentes conditions est difficile à comparer avec les performances des participants de l'**Expérience 4**, qui sont bien plus faibles. Nous supposons que ces disparités s'expliquent par le fait que le modèle n'a pas simulé les contraintes d'une réponse saccadique. Dans l'expérience comportementale, les participants ne pouvaient pas explorer librement les images. Ils devaient répondre en fixant le centre de l'écran, alors que le modèle avait un accès précis à chaque pixel des images. De plus, alors que les réponses saccadiques sont adaptées à l'étude du décours temporel du traitement visuel et des effets attentionnels, elles sont connues pour entraîner plus d'erreurs que les réponses manuelles (voir par exemple Bannerman, Milders et Sahraie, 2009 ; Kirchner et Thorpe, 2006). L'intégration de ces contraintes dans notre modèle ou l'utilisation de réponses manuelles plutôt que saccadiques dans l'**Expérience 4** aurait pu rendre les résultats plus comparables. Nous pourrions dans un premier temps envisager de développer une expérience laissant la possibilité aux participants d'explorer les images, comme une tâche de catégorisation d'émotions avec une réponse manuelle. En combinant ce type d'expérience avec un enregistrement des mouvements oculaires, nous pourrions directement comparer l'utilisation de l'information par le réseau et par les humains. Ainsi, il sera possible d'évaluer de manière plus précise si les cartes de saillance issues d'un CNN peuvent prédire les régions utilisées par les participants dans une tâche spécifique.

5.6 Conclusions

Pour conclure, ces travaux de thèse ont dans un premier temps permis d'éclaircir certains points théoriques concernant l'influence des expressions faciales dans la programmation des saccades oculaires. Ainsi, nous avons montré que les visages émotionnels peuvent être détectés et capturer le regard plus efficacement que les visages neutres, d'autant plus lorsqu'ils sont joyeux. Néanmoins, cet effet ne serait pas automatique, mais dépendant de la tâche. Nous suggérons que, en particulier dans les tâches qui induisent une réponse très rapide, comme les tâches de choix saccadique pour la détection de visages, les visages émotionnels ne vont pas capturer le regard plus que les visages neutres. Après l'étape de la détection des visages, les visages émotionnels vont pouvoir être détectés et capturer le regard plus efficacement que les visages neutres, même lorsque les expressions ne sont pas pertinentes pour la tâche. Des travaux futurs pourraient être envisagés pour mieux comprendre les conditions nécessaires à l'observation d'une capture du regard par les visages émotionnels. Ensuite, nous avons observé que, dès l'étape de la perception des visages, les expressions faciales peuvent influencer les points d'arrivée des saccades. En utilisant des images filtrées, nous avons aussi montré que la détection et le déclenchement de saccades vers des visages émotionnels reposaient sur le traitement des HFS plus que sur celui des BFS. Ainsi, contrairement à nos hypothèses, les visages apeurés, en BFS, n'attiraient pas particulièrement l'attention et regard. Cette hypothèse reposait sur l'idée de l'existence d'une voie sous-corticale, reliant le CS et l'amygdale, qui serait impliquée dans la détection de visages émotionnels, en particulier apeurés, et qui traiterait uniquement les BFS. Cependant, même au niveau neural, nos résultats n'ont pas montré d'effet des expressions faciales, ni en BFS ni en HFS, dans les régions qui constituent cette voie sous-corticale. Au regard de nos résultats, il semble que le traitement privilégié des visages émotionnels observé dans nos travaux puisse émerger d'un traitement cortical, impliquant la voie rétino-géniculo-striée et des aires corticales telles que l'OFA et la FFA. Néanmoins, nous n'excluons pas que certains biais méthodologiques, par exemple dans la construction de nos stimuli, puissent expliquer cette absence d'effet. Des travaux futurs pourraient être envisagés pour préciser plus exactement les réseaux cérébraux impliqués dans différents contextes (par exemple avec une présentation en vision périphérique ou avec différents filtrages).

Dans un second temps, ces travaux de thèse présentent également des contributions méthodologiques. Ils ont ainsi permis d'évaluer l'impact de l'égalisation du contraste d'images filtrées sur leur traitement. Cependant, les effets observés sont relativement limités. Nous avons montré que les saccades vers les visages émotionnels étaient effectuées plus rapidement en HFS qu'en BFS, mais seulement avec un contraste égalisé. Au niveau neural, les clusters occipito-temporaux impliqués dans le traitement des HFS étaient plus larges avec un contraste égalisé. Pour finir, nous avons au cours de ces travaux mis en avant l'intérêt de l'utilisation des réseaux de neurones artificiels en tant qu'outils pour étudier le comportement humain. D'abord, ils peuvent permettre la dissociation de facteurs perceptifs et émotionnels. Par exemple, nous avons montré que la meilleure

détection de visages émotionnels joyeux (en comparaison avec des visages émotionnels apeurés) peut s'expliquer par des facteurs physiques. Ensuite, ils peuvent être utilisés dans le cadre de la visualisation de l'importance des différentes parties d'une image pour une tâche précise. Dans nos travaux, nous avons montré que la région de la bouche était la plus utile dans une tâche de discrimination de visages neutres et émotionnels, et sa saillance était capable de prédire les performances des participants. Des travaux futurs pourraient étendre l'utilisation de ce modèle à l'étude des caractéristiques utiles dans des tâches diverses, et évaluer de manière plus précise à quel point les humains et le modèle utilisent l'information de manière similaire.

Appendices

Appendice A

Appendices au Chapitre 2

A.1 Analysis of the position of the eyes and mouth in face images depending on expressions

Here, we present an additional analysis conducted in order to compare the image positions of the eyes or mouth between expressions.More precisely, for each image, we manually drew a rectangular box surrounding the eyes and another surrounding the mouth, as presented in Figure A.1. We tried to select a rectangle that was as close as possible to the eyes or mouth but without touching the eyeball or lips, respectively. For each box, we computed the vertical midline (see purple lines in Figure A.1). We then computed two measurements : (1) the vertical distance between the center of the image (corresponding to a Y coordinate of 150 pixels) and the vertical midline of the eyes, and (2) the vertical distance between the center of the image and the vertical midline of the mouth. T-tests were performed on mean values for each expression to quantify the effect of the Emotional Facial Expression (Fearful, Happy, Neutral) on the position of the eyes and mouth. Measures are in degrees of visual angle to match the endpoint analysis that is presented in the manuscript. In our experimental setup, 1 degree corresponds to 25 pixels.

Results show that (1) vertical distance between the center of the image and the center of the eyes was smaller for neutral ($M \pm SD$: –0.076 ± 0.043 °) than for happy ($M \pm SD$: –0.05 ± 0.052 °, $t(1,59) = 2.97$, $p = .004$, $d = 0.54$) or fearful ($M \pm SD$: –0.024 ± 0.05 °, $t(1,59) = 6.13$, $p < .001$, $d = 1.12$) faces, and for happy than for fearful faces ($t(1,59) = –2.84$, $p = .005$, $d = 0.52$), and that (2) vertical distance between the center of the image and the center of the mouth was smaller for fearful ($M \pm SD$: –3.60 ± 0.19 °) than for happy ($M \pm SD$: –3.38 ± 0.14 °, $t(1,59) = 7.31$, $p < .001$, $d = 1.34$) or neutral ($M \pm SD$: –3.49 ± 0.14 °, $t(1,59) = –3.8$, $p < .001$, $d = 0.69$) faces, and for neutral than for happy faces ($t(1,59) = 4.22$, $p < .001$, $d = 0.77$).

In conclusion, based on the regions that we manually selected, there were significant differences concerning the positions of the eyes and mouth. The differences in the positions of the eyes were small (with the greatest mean difference being 1.25 pixels). The difference was larger when we compared the mean position of the mouth for the three emotions (with the greatest mean difference being 5.5 pixels).

A.2 The effect of the target location

Here, we present an additional analysis conducted in order to compare participants' performance and saccade endpoints depending on the target location (i.e., left or right

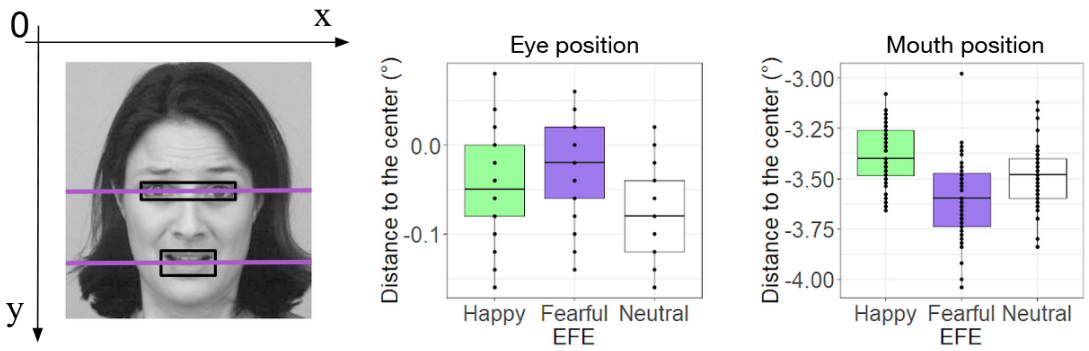

Figure A.1 – Example of rectangular boxes surrounding the eyes and mouth with their vertical midlines (purple lines ; left), and boxplots for the vertical distances between the image center and the eyes (middle), and between the image center and the mouth (right). A NonEGative distance means that the feature is below the center.

visualhemifield ; Figure A.2). For each experiment, we performed (1) a repeated measures ANOVA on the mean accuracy, mean latency, and mean distance to the center, with the Target (Face, Vehicle for Experiment 1 ; Emotional, Neutral for Experiment 2) and the Location (Left, Right) as between-subject factors to test for the effect of the Location and for an interaction between Target and Location. Then, for each Target, we performed (2) a repeated measures ANOVA on the mean accuracy, latency, and distance to the center of one target condition (face condition for Experiment 1, emotional condition for Experiment 2), with the EFE (Happy, Fearful, Neutral for Experiment 1 ; Happy, Fearful for Experiment 2) and Location (Left, Right) as between-subject factors to test for an interaction between the EFE of the target and the Location. This procedure was repeated for the other target condition (vehicle condition for Experiment 1, neutral condition for Experiment 2) to test for an interaction between distractor Location and EFE. When required, paired samples t-tests were used for pairwise comparisons. Neither main effects of the Target and the EFE (which are reported in the core paper) nor nonsignificant effects are reported.

Experiment 1 : Face versus vehicle

Accuracy : A repeated measures ANOVA performed on mean accuracy when the target was a face indicated a marginal effect of Location ($F(1,60) = 3.71$, $p = .059$, $\eta_p^2 = 0.029$). Saccades tended to be more accurate when the face appeared in the left ($M \pm SD$: .89 ± .087) than the right ($M \pm SD$: .87 ± .099) hemifield.

Latency : A repeated measures ANOVA performed on mean latency indicated an interaction between the Target and Location ($F(1,60) = 15.1$, $p < .001$, $\eta_p^2 = 0.2$). When the target was a face, saccades were elicited faster when it was presented in the left ($M \pm SD$: 178 ± 21.2 ms) than in the right ($M \pm SD$: 174 ± 23 ms ; $p <$ .001) visual hemifield. *Endpoints :* A repeated measures ANOVA performed on mean distance to the center indicated a significant effect of Location ($F(1,60) = 7.09$, $p <$ .001, $\eta_p^2 = 0.11$). Saccades landed lower in the left ($M \pm SD$: –0.34 ± 0.55) than in the right ($M \pm SD$: –0.23 ± 0.48) hemifield.

Experiment 2 : Emotional versus neutral face

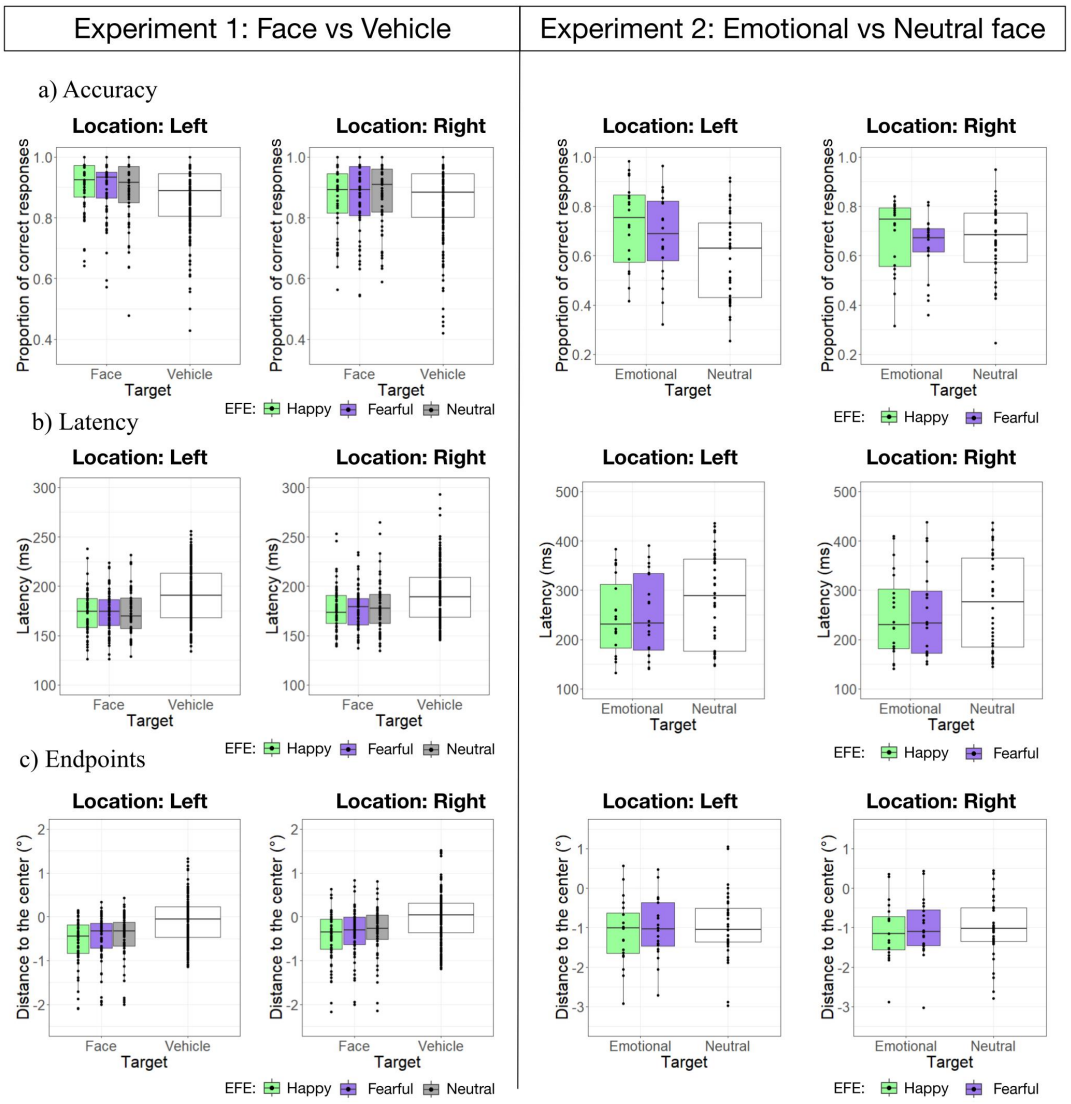

Figure A.2 – Boxplots for (a) mean proportion of correct responses, (b) mean latency (in ms), and (c) mean distance to the image center (in degrees of visual angle) for correct saccades according to the Target, Location of targets, and Emotional Facial Expression of targets, for Experiment 1 (left) and Experiment 2 (right).

Accuracy : A repeated measures ANOVA performed on mean accuracy indicated a significant interaction between Target and Location ($F(1,19) = 6$, $p < .001$, $\eta_p^2 = 0.024$). The difference between emotional and neutral faces was only significant for the left hemifield (accuracy for emotional faces : $M \pm SD$: $.7 \pm .16$; accuracy for neutral faces : $M \pm SD$: $.59 \pm .18$; $p < .001$).

Latency and endpoints : A repeated measures ANOVA performed on mean latency or mean distance to the center indicated no significant effect of Location, nor any interaction with other factors.

Appendice B

Appendices au Chapitre 3

B.1 Saliency toolbox interface

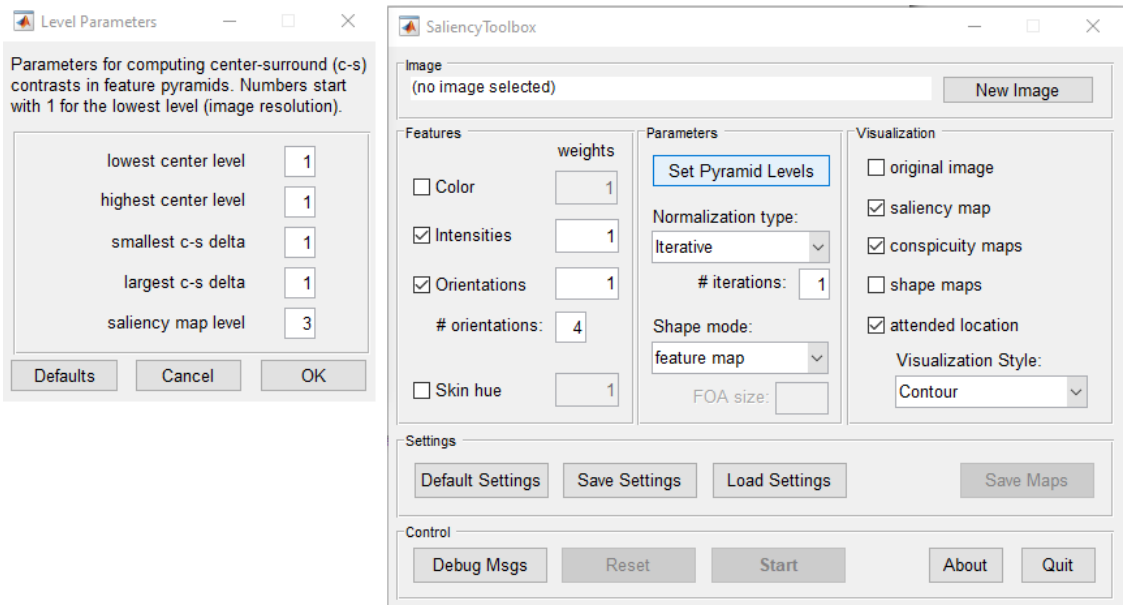

Figure B.1 – Overview of the Saliency toolbox interface (Walther and Koch, 2006) with the parameters used to generate the saliency maps in this study.

B.2 Detailed results for human performance

Paired-samples t-test revealed that the accuracy was overall higher when the target was emotional ($M \pm SD$: .65 ± .12) than neutral ($M \pm SD$: .59 ± .12; $t(77) = 8.17$, $p < .001$, $d = 0.92$).

When the target was the emotional face, repeated measures ANOVA revealed a main effect of the EFE of the target ($F(1,77) = 8.44$, $p = .005$, $\eta_p^2 = 0.099$), and of the Spatial Frequency ($F(2,154) = 6.05$, p = .003, $\eta_p^2 = 0.073$). The accuracy was higher when the target was happy ($M \pm SD$: .66 ± .13) than fearful ($M \pm SD$: .64 ± .12), and for images in BSF ($M \pm SD$: .66 ± .13; $p < .001$) and HSF ($M \pm SD$: .65 ± .12; $p = .036$) than LSF ($M \pm SD$: .63 ± .13). There was also an interaction between the Spatial Frequency and the EFE ($F(2,154) = 8.44$, $p = .005$, $\eta_p^2 = 0.086$). The difference between happy and fearful faces was only significant in HSF ($M \pm SD$ for happy faces :

.68 ± .15 ; $M \pm SD$ for fearful faces : .62 ± .12), and the difference between HSF and LSF was only significant with a happy face ($M \pm SD$ for LSF : .62 ± .14).

When the target was the neutral face, repeated measures ANOVA revealed a marginal effect of the EFE of the distractor ($F(1,77) = 3.83$, $p = .053$, $\eta_p^2 = 0.047$). The accuracy was higher when the distractor was happy ($M \pm SD$: .6 ± .13) than fearful ($M \pm SD$: .58 ± .12). There was also an interaction between the Spatial Frequency and the EFE ($F(2,154) = 4.26$, $p = .016$, $\eta_p^2 = 0.052$). The difference between happy and fearful faces was only significant in HSF ($M \pm SD$ for happy faces : .62 ± .15 ; $M \pm SD$ for fearful faces : .57 ± .14), and the accuracy was marginally higher in HSF than LSF only with a happy face ($M \pm SD$ for LSF : .59 ± .14). Overall, with these behavioral results we notably observed that emotional faces are more attractive than neutral faces and that their detection is easier for HSF than LSF pairs, with happy emotional faces.

B.3 Detailed results for saccade endpoints

Paired-samples t-test revealed that the endpoints went lower when the target was emotional ($M \pm SD$: -1.28 ± 0.71) than neutral ($M \pm SD$: -1.11 ± 0.73 ; $t(77) =$ -4, $p < .001$, $d = 0.45$).

When the target was the emotional face, repeated measures ANOVA revealed a main effect of the EFE of the target ($F(1,77) = 13.2$, $p < .001$, $\eta_p^2 = 0.15$), and of the Spatial Frequency ($F(2,154) = 7.41$, $p = .001$, $\eta_p^2 = 0.088$). The endpoints were lower when the target was happy ($M \pm SD$: -1.31 ± 0.71) than fearful ($M \pm SD$: -1.27 ± 0.71), and for images in LSF ($M \pm SD$: -1.41 ± 0.78) than HSF ($M \pm SD$: -1.26 ± 0.76 ; $p = .05$) or BSF ($M \pm SD$: -1.18 ± 0.79 ; $p < .001$). There was also an interaction between the Spatial Frequency and the EFE ($F(2,154) = 10.4$, $p < .001$, $\eta_p^2 = 0.12$). The difference between happy and fearful targets was only significant in HSF ($M \pm SD$ for happy faces : -1.3 ± 0.75 ; $M \pm SD$ for fearful faces : -1.23 ± 0.78) and BSF ($M \pm SD$ for happy faces : -1.22 ± 0.8 ; $M \pm SD$ for fearful faces : -1.14 ± 0.79). The difference between HSF and LSF was only significant, although marginally, with a fearful face ($M \pm SD$ for LSF : -1.43 ± 0.79 ; $p = .08$).

When the target was the neutral face, repeated measures ANOVA revealed only an effect of the Spatial Frequency ($F(2,154) = 8.34$, $p < .001$, $\eta_p^2 = .098$). The endpoints went lower for images in LSF ($M \pm SD$: -1.24 ± 0.79) than HSF ($M \pm SD$: -1.08 ± 0.77 ; $p < .001$) or BSF ($M \pm SD$: -1.03 ± 0.79 ; $p < .001$). For a visual representation, Figure 7 (a) displays heat maps computed from saccade endpoints convolved with a small 2D gaussian in each conditions for all participants.

B.4 Detailed results for RMS_{CNN}

Paired-samples t-test revealed that the RMS_{CNN} was overall higher for emotional ($M \pm SD$: .19 ± .022) than neutral faces ($M \pm SD$: .084 ± .001 ; $t(199) = 72.6$, $p < .001$, $d = 5.1$).

For emotional faces, repeated measures ANOVA revealed a main effect of the EFE ($F(1,199) = 282$, $p < .001$, $\eta_p^2 = 0.017$), and of the Spatial Frequency ($F(2,398) = 159$, $p < .001$, $\eta_p^2 = 0.055$). The RMS_{CNN} was higher for happy ($M \pm SD : .19 \pm .024$) than fearful ($M \pm SD : .18 \pm .021$) faces, and for images in BSF ($M \pm SD : .19 \pm .025$; $p < .001$) and HSF ($M \pm SD : .19 \pm .024$; $p < .001$) than LSF ($M \pm SD : .17 \pm .025$), as well as for images in BSF than HSF ($p < .001$). There was also an interaction between the Spatial Frequency and the EFE ($F(2,398) = 102$, $p < .001$, $\eta_p^2 = 0.004$). With a happy face, the RMS_{CNN} was higher in BSF ($M \pm SD : .205 \pm .03$) than HSF ($M \pm SD : .196 \pm .027$), whereas this was the opposite with a fearful face ($M \pm SD$ for HSF faces : $.19 \pm .023$; $M \pm SD$ for BSF faces : $.0.18 \pm .022$; $p < .001$).

For neutral faces, repeated measures ANOVA revealed a main effect of the EFE ($F(1,199) = 12.1$, $p = .015$, $\eta_p^2 = 0.00001$), and of the Spatial Frequency ($F(2,398) = 4.23$, $p < .001$, $\eta_p^2 = 0.006$). The RMS_{CNN} of neutral faces was higher with a happy ($M \pm SD : .0837 \pm .008$) than a fearful ($M \pm SD : .0836 \pm .007$) distractor, and for images in BSF ($M \pm SD : .084 \pm .009$) than HSF ($M \pm SD : .083 \pm .009$; $p < .001$) and LSF ($M \pm SD : .083 \pm .01$; $p = .004$). There was also an interaction between the Spatial Frequency and the EFE ($F(2,398) = 7.55$, $p < .001$, $\eta_p^2 = 0.000009$). The difference between happy and fearful faces was only significant for BSF images ($M \pm SD$ for happy faces : $.0847 \pm .0009$; $M \pm SD$ for fearful faces : $.0.845 \pm .009$; $p < .001$). Note that, due to their small effect size, the main effect of the EFE and the interactions between the EFE and the Spatial Frequency for neutral faces are not mentioned in the result section.

B.5 Detailed results for $\text{RMS}_{Bottom-up}$

Welch Two Sample t-test revealed that the $\text{RMS}_{Bottom-up}$ ($M \pm SD : .075 \pm .055$) was overall lower than the RMS_{CNN} ($M \pm SD : .13 \pm .062$); $t(452.7) = 18.41$, $p < .001$, $d = 1.06$).

Welch Two Sample t-test revealed that the $\text{RMS}_{Bottom-up}$ was overall higher for emotional ($M \pm SD : .11 \pm .07$) than neutral faces ($M \pm SD : .042 \pm .025$); $t(398) = 12.8$, $p < .001$, $d = 1.28$). For emotional faces, two-way ANOVA revealed a main effect of the Spatial Frequency ($F(2,194) = 17.1$, $p < .001$, $\eta_p^2 = 0.16$). The $\text{RMS}_{Bottom-up}$ was higher for images in BSF ($M \pm SD : .094 \pm .071$) and HSF ($M \pm SD : .13 \pm .06$) than LSF ($M \pm SD : .077 \pm .067$; $p < .001$). There was also an interaction between the Spatial Frequency and the EFE ($F(2,194) = 10.5$, $p < .001$, $\eta_p^2 = 0.11$). In HSF, the $\text{RMS}_{Bottom-up}$ was higher for happy ($M \pm SD : .15 \pm .047$) than fearful faces ($M \pm SD : .11 \pm .064$; $p = .04$), whereas this was the opposite in LSF ($M \pm SD$ for happy faces : $.054 \pm .033$; $M \pm SD$ for fearful faces : $.0.1 \pm .083$; $p = .01$). Also, the higher saliency in HSF and BSF than LSF was only significant with a happy face ($M \pm SD$ for BSF faces : $.13 \pm .051$; $p < .001$).

For neutral faces, two-way ANOVA revealed a main effect of the Spatial Frequency ($F(2,194) = 20.12$, $p < .001$, $\eta_p^2 = 0.088$). The $\text{RMS}_{Bottom-up}$ was higher for images in BSF ($M \pm SD : .062 \pm .029$) than in HSF ($M \pm SD : .039 \pm .018$) or LSF ($M \pm SD : .034 \pm .024$; $p < .001$).

Appendice C

Appendices au Chapitre 4

C.1 Description du scan utilisé pour la localisation fonctionnelle

Le scan fonctionnel utilisé pour localiser les aires sélectives aux visages était composé de 16 blocs de 15 secondes chacun. Ces blocs correspondaient à la présentation de visages neutres (4 blocs), de visages émotionnels (4 blocs), d'objets (4 blocs) ou à une période de repos (4 blocs). Ainsi, il y avait 12 blocs de tâche et 4 blocs de repos. Dans les blocs de tâche, les images étaient présentées sous différents niveaux de contrastes. Afin de maintenir l'attention des participants sur les images, les participants ont effectué une tâche de détection de répétition. Ils avaient pour instruction d'appuyer sur un bouton à chaque fois qu'ils voyaient deux stimuli identiques. Chaque stimulus était présenté pendant 600 ms, avec un intervalle interstimulus de 400 ms. Dans chaque bloc, deux répétitions avaient lieu. La Figure C.1 présente une visualisation du déroulement d'un scan. Comme pour notre expérience d'intérêt, les images fonctionnelles ont été acquises par l'intermédiaire de séquences pondérées en T2*, en écho de gradient rapide (echo planar imaging, EPI), avec les paramètres suivants : TR/TE = 2500/30 ms, angle de basculement = 80°, matrice d'acquisition = 80 x 157, champ de vue = 120 x 240 x 72, 48 coupes transversales, épaisseur des coupes = 1.35 mm, taille des voxels = 1.5 x 1.5 x 1.35 mm.

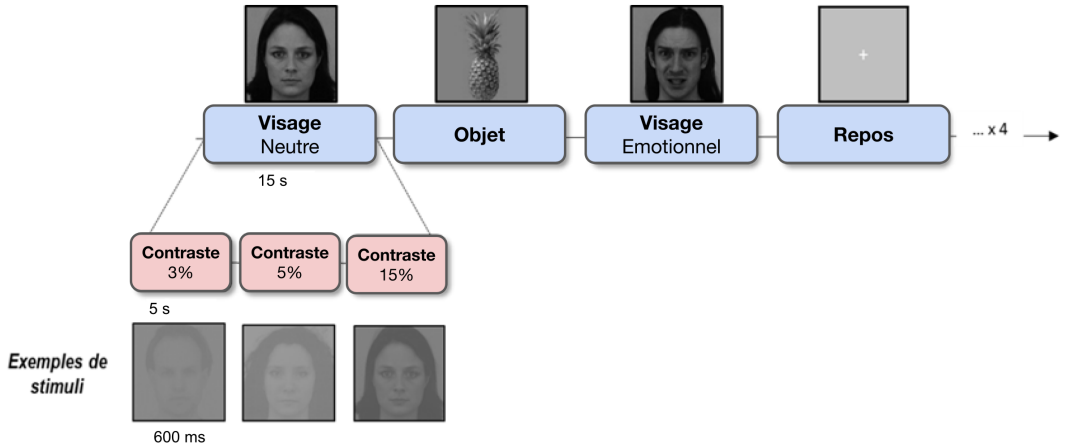

Figure C.1 – Procédure utilisée pour la localisation des aires des visages.

C.2 Localisation de la FFA et de l'OFA pour chaque participant

	FFA		OFA	
Participant	Droit	Gauche	Droit	Gauche
4	x*	x	x	
5		x		x
6		x		x
8	x*	x	x*	x
9	x		x	
12	x		x*	x
13	x	x*		
19	x			x
22	x		x	
23				
25			x*	x
30	x	x*	x	
31	x*	x	x*	x
33	x		x*	x
34	x*	x	x*	x
35	x	x*	x*	x
36	x		x	
37	x			
38	x		x	
39	x*	x	x*	x
40	x	x*		
41	x		x	x*
42	x	x*	x	x*
44				
45			x	

Table C.1 – Localisation de la FFA et l'OFA pour chaque participant. Un x indique que la région a été localisée. Un marquage * indique l'hémisphère qui a été retenu lorsque la région était localisée dans les deux hémisphères. La FFA a été localisée des dans les deux hémisphères chez 10 participants, dans l'hémisphère doit seulement chez 8 participants et dans l'hémisphère gauche seulement chez 2 participants. L'OFA a été localisée des dans les deux hémisphères chez 11 participants (incluant le participant 22, pour lequel un cluster a été localisé mais il s'étendait sur les deux hémisphères), dans l'hémisphère doit seulement chez 6 participants et dans l'hémisphère gauche seulement chez 3 participants.

C.3 Clusters obtenus pour l'effet des émotions et des fréquences spatiales

Apeurée >Neutre						
k	**Label (AAL)**	**x**	**y**	**z**	**Pic (T)**	**Pic (Z)**
BFS_NonEG						
17	Temporal_Sup_L	-50	-7	15	3.83	3.76
12	Fusiform_L	-35	-57	-8	3.7	3.63
6		-27	-55	-9	3.5	3.44
6	Putamen_L	-26	7	3	3.51	3.45
BFS_EG						
23	Putamen_R	29	-13	-2	4.49	4.37
67	Thal_VL_L	-15	-13	7	4.24	4.14
7	Temporal_Inf_R	56	-22	-19	4.13	4.03
27	Fusiform_L	-24	-37	-15	4.04	3.95
14	Occipital_Sup_L	-21	-82	23	3.99	3.9
12	Hippocampus_L	-30	-22	-12	3.9	3.82
33	Insula_R	32	31	6	3.87	3.79
8	Parahippocampal_R	24	-22	-20	3.76	3.69
7	Hippocampus_R	35	-15	-21	3.73	3.66
12	OFCpost_R	39	19	-19	3.52	3.46
HFS_NonEG						
16	Fusiform_R	41	-42	19	4.13	4.04
9		38	-78	-17	3.87	3.79
38	Thal_MDm_R	3	-15	-2	3.95	3.87
6	Putamen_R	32	7	8	3.85	3.78
14	Temporal_Mid_R	-44	-72	19	3.52	3.46
19	Temporal_Inf_R	51	-67	-7	3.38	3.33
HFS_EG						
28	Cyngulate_Post_L	-5	-49	29	4.23	4.13
139	Fusiform_R	45	-49	-23	4.17	4.07
20	Frontal_Inf_Orb_2_R	33	25	-12	4.13	4.03
13	Occipital_Mid_R	38	-72	10	3.97	3.88
16	Temporal_Mid_R	44	-64	12	3.76	3.69
9	Frontal_Inf_Orb_2_L	-42	20	-8	3.67	3.6
Global						
348	Fusiform_R	45	-48	-21	5	4.84
7		38	-76	19	3.43	3.38
35	Fusiform_L	-30	-40	-19	3.43	3.38
6		-36	-57	-8	3.85	3.77
6	Amygdala_L	-29	-1	-24	3.44	3.39
61	Temporal_Mid_R	44	-51	3	4.51	4.39
	Temporal_Inf_L	-45	-46	-24	3.89	3.82
10	Temporal_Mid_L	-51	-72	3	4.21	4.11
17	Temporal_Sup_R	54	-4	-15	3.67	3.6
34	OFCpost_R	41	20	-17	4.14	4.05
10	Insula_L	-42	7	-9	3.89	3.81
49	Occipital_Inf_L	-47	-70	-12	3.83	3.75
17		-38	-82	-11	3.81	3.74

Table C.2 – Coordonnées des pics d'activation et labels (AAL) associés à la perception d'un visage apeuré (comparativement à un visage neutre), pour chaque condition de Fréquences Spatiales (BFS, HFS) et de Contraste (NonEG, EG), et de manière globale, indépendamment des fréquences spatiales et du contraste.

BFS >HFS						
k	**Label (AAL)**	**x**	**y**	**z**	**Pic (T)**	**Pic (Z)**
Apeurée_NonEG						
9	Acc_sup_R	11	25	22	3.7	3.63
5	Putamen_L	-26	8	3	3.65	3.59
10		-20	5	-8	3.59	3.52
6	Frontal_Mid_2_R	35	38	10	3.67	3.61
6	Temporal_Mid_R	-39	-64	16	3.32	3.27
Apeurée_EG						
48	Caudate_L	-12	14	-4	4.52	4.4
12	Precuneus_R	26	-48	10	3.7	3.64
Neutre_NonEG						
81	Occipital_Mid_L	-41	-78	10	4.62	4.49
24	Temporal_Mid_R	48	-73	12	4.23	4.13
10		54	-25	-8	3.74	3.67
17	Insula_L	-29	13	-7	4.09	4
13	Acc_sup_R	18	35	19	3.68	3.61
90	Thal_MDm_R	2	-15	-2	4.24	4.14
38	Temporal_Mid_L	-42	-67	8	3.94	3.86
5	Insula_R	42	-1	-8	3.47	3.41
32	Frontal_Mid_2_L	-27	55	15	3.66	3.59
18	Lingual_R	11	-34	-9	3.52	3.46
6	Calcarine_R	20	-72	18	3.51	3.45
Neutre_EG						
15	Insula_L	-26	32	11	3.88	3.8
50	Hippocampus_R	29	-43	4	4.78	4.64
9	Temporal_Mid_R	45	-63	6	3.8	3.73
7	Precuneus_R	26	-51	14	3.57	3.51
6	Temporal_Mid_L	-36	-64	14	3.44	3.39
5	Calcarine_R	26	-72	4	3.4	3.35
Global						
31	Lingual_L	-23	-75	3	4.76	4.62
32	Calcarine_R	23	-76	3	4.12	4.03
58	Temporal_Mid_R	44	-72	4	4.03	3.95
59	Temporal_Mid_L	-39	-64	14	4.31	4.21
37	Occipital_Sup_R	24	-67	18	4.26	4.16
115	Occipital_Mid_L	-39	-75	10	4.23	4.13
119	Precuneus_R	29	-43	6	4.94	4.79

Table C.3 – Coordonnées des pics d'activation et labels (AAL) associés à la perception d'un visage en BFS (comparativement à un visage en HFS), pour chaque condition d'EFE (Apeurée, Neutre) et de Contraste (NonEG, EG), et de manière globale, indépendamment de l'EFE et du contraste.

BFS <HFS						
k	**Label (AAL)**	**x**	**y**	**z**	**Pic (T)**	**Pic (Z)**
Apeurée_NonEG						
1817	Lingual_R	26	-87	-11	6.56	6.23
2923	Fusiform_L	-12	-91	-11	6.46	6.13
9		-42	-67	-21	3.66	3.6
581	Fusiform_R	30	-57	-19	5.52	5.31
85	Thal_PuM_R	20	-30	3	4.45	4.33
33	Temporal_Sup_R	56	-33	6	4.4	4.29
20	Temporal_Inf_R	47	-66	-7	4.07	3.98
9		45	-39	-21	3.73	3.66
7	Insula_R	39	2	15	3.42	3.36
Apeurée_EG						
5501	Fusiform_R	15	-97	3	9.41	Inf
8		42	-39	-21	3.69	3.63
5111	Fusiform_L	-14	-91	-12	8.91	Inf
5		-18	-34	-17	3.47	3.41
14	Temporal_Mid_R	50	-19	-13	3.67	3.6
9	Temporal_Pole_Sup_R	41	11	-23	3.64	3.58
8	Temporal_Pole_Sup_L	-27	8	-23	3.43	3.47
8	Insula_R	27	16	-21	3.59	3.53
9	Insula_L	-27	7	15	3.58	3.52
17	Lingual_R	21	-52	-11	3.48	3.43
Neutre_NonEG						
1407	Occipital_Inf_L	-14	-90	-13	7.38	6.92
1685	Lingual_R	14	-94	-4	5.82	5.58
99	Hippocampus_R	41	-27	-13	4.89	4.75
125	Fusiform_L	-29	-55	-11	4.49	4.37
24		-29	-73	-8	3.77	3.7
65	Lingual_L	-24	-64	-12	4.27	4.16
90	Fusiform_R	33	-52	-17	4.09	4
25		32	-60	-5	3.62	3.55
24	Temporal_Inf_L	-53	-49	-8	4.16	4.07
15	Calcarine_R	26	-60	11	3.72	3.65
Neutre_EG						
5208	Fusiform_R	15	-99	2	10	Inf
4814	Fusiform_L	-14	-91	-12	9.42	Inf
56		-26	-39	-15	3.78	3.71
15	Amygdala_R	26	-3	-19	3.86	3.78
13	Amygdala_L	-21	-3	-19	3.39	3.34
54	Frontal_Inf_Orb_2_R	39	34	-13	3.89	3.81
78	Thal_VL_L	-17	-12	8	4.18	4.09
6	Putamen_R	30	-16	0	3.42	3.37
Global						
8424	Fusiform_R	15	-97	2	13.59	Inf
7614	Fusiform_L	-14	-91	-12	14.74	Inf
22		-33	-31	-24	3.65	3.58
12	Amygdala_L	-18	-6	-16	3.69	3.63
210	Hippocampus_R*	17	-6	-16	4.39	4.28
213	Cingulate_Post_L	-2	-36	33	5.12	4.95
31	OFCpost_L	-29	32	-20	4.53	4.41
15	Insula_R	26	16	-23	3.94	3.86
21	ParaHippocampal_R	20	7	-21	3.86	3.79
36	Thal_LGN_R	18	-28	0	3.75	3.67
16	Temporal_Mid_R	57	-12	-19	4.2	4.1
14	Temporal_Pole_Sup_R	45	8	-23	3.76	3.69
16	Temporal_Sup_R	56	-33	6	4.09	4

Table C.4 – Coordonnées des pics d'activation et labels (AAL) associés à la perception d'un visage en HFS (comparativement à un visage en BFS), pour chaque condition d'EFE (Apeurée, Neutre) et de Contraste (NonEG, EG), et de manière globale, indépendamment de l'EFE et du contraste. *Le cluster labellisé Hippocampe droit comprend une partie de l'amygdale droite (34% du cluster, contre 47% pour l'hippocampe).

C.4 Cartes des activations obtenues pour l'effet des émotions et des fréquences spatiales

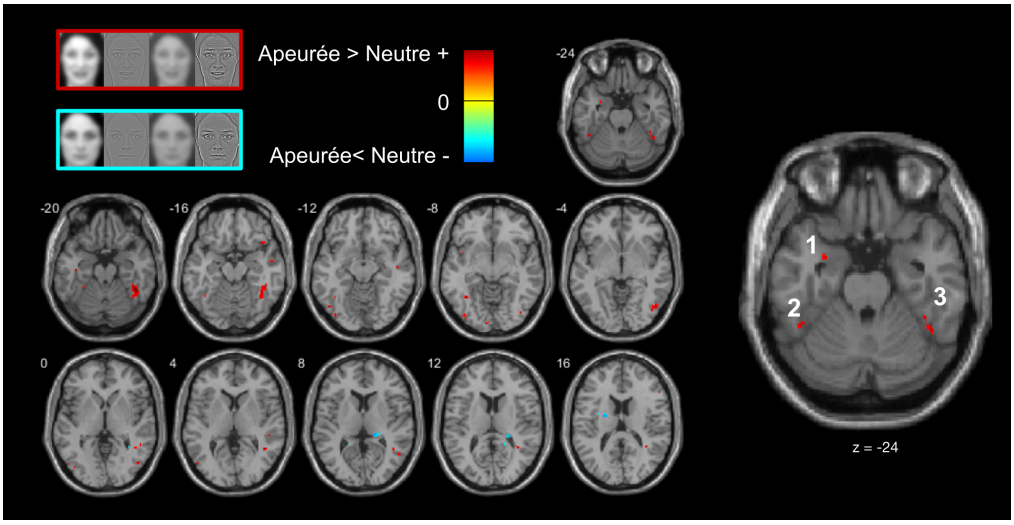

Figure C.2 – Aperçu des cartes des activations obtenues pour le contraste [Apeurée-Neutre], avec un seuil de significativité p <.001 et un nombre minimal de voxels par cluster $k = 5$. Les valeurs positives, en rouges, correspondent aux régions qui s'activent significativement plus pour les visages apeurés que neutres. Zoom sur les régions de l'amygdale (1), et des gyrus fusiformes gauche (2) et droit (3).

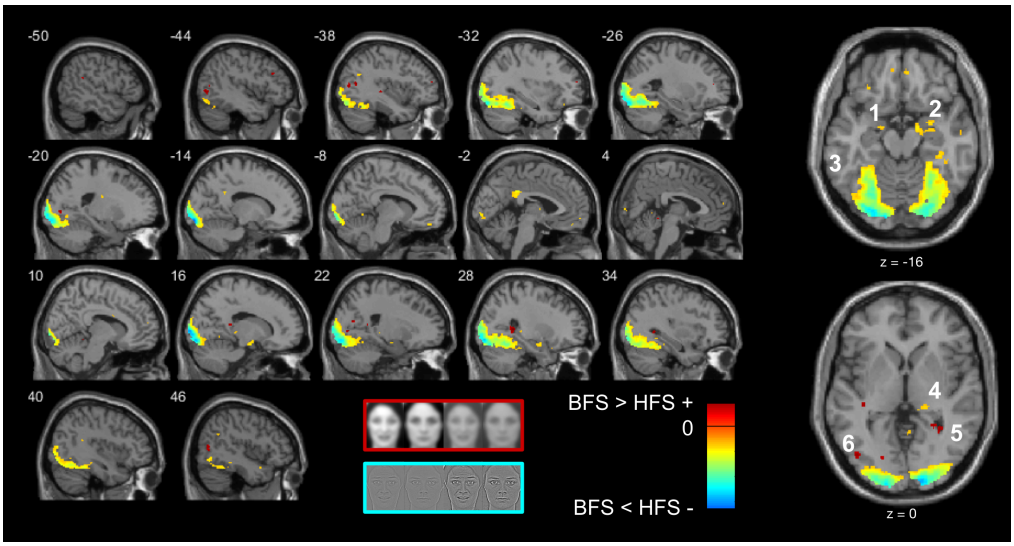

Figure C.3 – Aperçu des cartes des activations obtenues pour le contraste [BFS-HFS], avec un seuil de significativité p <.001 et un nombre minimal de voxels par cluster $k = 5$. Les valeurs positives, en rouges, correspondent aux régions qui s'activent significativement plus pour les visages en BFS qu'en HFS. Les valeurs négatives, en bleu, correspondent aux régions qui s'activent significativement plus pour les visages en HFS qu'en BFS. Zoom sur les régions de l'amygdale gauche (1) et droite (2), du gyrus fusiforme gauche (3), du thalamus incluant le pulvinar (4), du gyrus temporal moyen droit (5) et du gyrus occipital moyen gauche (6).

Appendice D

Ressources partagées

Les données expérimentales ainsi que la plupart des programmes associés aux 3 articles présentés dans ce travail de thèse sont disponibles sur l'*Open Science Framework*.

Article 1 : https://osf.io/bjmcy.

Article 2 : https://osf.io/hyr52.

Article 3 : https://osf.io/vq3jy.

Références

Adams Jr, R. B., Gordon, H. L., Baird, A. A., Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science*, *300*(5625), 1536-1536.

Adams Jr, R. B., Hess, U., & Kleck, R. E. (2015). The intersection of gender-related facial appearance and facial displays of emotion. *Emotion Review*, *7*(1), 5-13.

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 21-62.

Adolphs, R. (2008). Fear, faces, and the human amygdala. *Current opinion in neurobiology*, *18*(2), 166-172.

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., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669-672.

Aguado, L., Garcia-Gutierrez, A., Castañeda, E., & Saugar, C. (2007). Effects of prime task on affective priming by facial expressions of emotion. *The Spanish Journal of Psychology*, *10*(2), 209-217.

Ahs, F., Davis, C. F., Gorka, A. X., & Hariri, A. R. (2014). Feature-based representations of emotional facial expressions in the human amygdala. *Social cognitive and affective neuroscience*, *9*(9), 1372-1378.

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. (2003). Neural correlates of the automatic processing of threat facial signals. *Journal of Neuroscience*, *23*(13), 5627-5633.

Andino, S. L. G., de Peralta Menendez, R. G., Khateb, A., Landis, T., & Pegna, A. J. (2009). Electrophysiological correlates of affective blindsight. *NeuroImage*, *44*(2), 581-589.

Ariga, A., & Arihara, K. (2018). Attentional capture by spatiotemporally task-irrelevant faces : supportive evidence for Sato and Kawahara (2015). *Psychological research*, *82*(5), 859-865.

Asghar, A. U., Chiu, Y.-C., Hallam, G., Liu, S., Mole, H., Wright, H., & Young, A. W. (2008). An amygdala response to fearful faces with covered eyes. *Neuropsychologia*, *46*(9), 2364-2370.

Awasthi, B., Friedman, J., & Williams, M. A. (2011). Faster, stronger, lateralized: low spatial frequency information supports face processing. *Neuropsychologia*, *49*(13), 3583-3590.

Bach, D. R., Hurlemann, R., & Dolan, R. J. (2015). Impaired threat prioritisation after selective bilateral amygdala lesions. *cortex*, *63*, 206-213.

Badcock, J. C., Whitworth, F. A., Badcock, D. R., & Lovegrove, W. J. (1990). Low-frequency filtering and the processing of local—global stimuli. *Perception*, *19*(5), 617-629.

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual cognition*, *17*(6-7), 1185-1204.

Bannerman, R. L., Hibbard, P. B., Chalmers, K., & Sahraie, A. (2012a). Saccadic latency is modulated by emotional content of spatially filtered face stimuli. *Emotion*, *12*(6), 1384-1392.

Bannerman, R. L., Hibbard, P. B., Chalmers, K., & Sahraie, A. (2012b). Saccadic latency is modulated by emotional content of spatially filtered face stimuli. *Emotion*, *12*(6), 1384.

Bannerman, R. L., Milders, M., de Gelder, B., & Sahraie, A. (2009). Orienting to threat: faster localization of fearful facial expressions and body postures revealed by saccadic eye movements. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1662), 1635-1641.

Bannerman, R. L., Milders, M., & Sahraie, A. (2009). Processing emotional stimuli: comparison of saccadic and manual choice-reaction times. *Cognition & Emotion*, *23*(5), 930-954.

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., & Neta, M. (2007). Visual elements of subjective preference modulate amygdala activation. *Neuropsychologia*, *45*(10), 2191-2200.

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM) : Two cases for evolutionary psychology. *Joint Attention : Its Origins and Role in Development*, 41-59.

Barrett, L. F. (2011). Was Darwin wrong about emotional expressions ? *Current Directions in Psychological Science*, *20*(6), 400-406.

Barron, D. S., Eickhoff, S. B., Clos, M., & Fox, P. T. (2015). Human pulvinar functional organization and connectivity. *Human brain mapping*, *36*(7), 2417-2431.

Barten, P. G. (2003). Formula for the contrast sensitivity of the human eye. *Image Quality and System Performance*, *5294*, 231-238.

Bayle, D. J., Henaff, M.-A., & Krolak-Salmon, P. (2009). Unconsciously perceived fear in peripheral vision alerts the limbic system : a MEG study. *PLoS One*, *4*(12), e8207.

Bayle, D. J., Schoendorff, B., Hénaff, M.-A., & Krolak-Salmon, P. (2011). Emotional facial expression detection in the peripheral visual field. *PloS one*, *6*(6), e21584.

Bayle, D. J., & Taylor, M. J. (2010). Attention inhibition of early cortical activation to fearful faces. *Brain Research*, *1313*, 113-123.

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. *Cognition & emotion*, *28*(3), 416-432.

Becker, D. V., Anderson, U. S., Mortensen, C. R., Neufeld, S. L., & Neel, R. (2011). The face in the crowd effect unconfounded: happy faces, not angry faces, are more efficiently detected in single- and multiple-target visual search tasks. *Journal of Experimental Psychology: General*, *140*(4), 637-659.

Becker, S. I., Dutt, N., Vromen, J. M., & Horstmann, G. (2017). The capture of attention and gaze in the search for emotional photographic faces. *Visual Cognition*, *25*(1-3), 241-261.

Belopolsky, A. V. (2015). Common priority map for selection history, reward and emotion in the oculomotor system. *Perception*, *44*(8), 920-933.

Berboth, S., & Morawetz, C. (2021). Amygdala-prefrontal connectivity during emotion regulation : A meta-analysis of psychophysiological interactions. *Neuropsychologia*, *153*, 107767.

Bernstein, M., Erez, Y., Blank, I., & Yovel, G. (2018). An integrated neural framework for dynamic and static face processing. *Scientific reports*, *8*(1), 1-10.

Bertini, C., Cecere, R., & Ladavas, E. (2019). Unseen fearful faces facilitate visual discrimination in the intact field. *Neuropsychologia*, *128*, 58-64.

Bisley, J. W., & Mirpour, K. (2019). The neural instantiation of a priority map. *Current Opinion in Psychology*, *29*, 108-112.

Bisti, S., & Sireteanu, R. C. (1976). Sensitivity to spatial frequency and contrast of visual cells in the cat superior colliculus. *Vision research*, *16*(3), 247-251.

Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., & Gosselin, F. (2017). Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion*, *17*(7), 1107.

Blanchard, D. C., & Blanchard, R. J. (1972). Innate and conditioned reactions to threat in rats with amygdaloid lesions. *Journal of comparative and physiological psychology*, *81*(2), 281.

Bocanegra, B. R., & Zeelenberg, R. (2009). Emotion improves and impairs early vision. *Psychological science*, *20*(6), 707-713.

Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion recognition : The role of featural and configural face information. *Quarterly Journal of Experimental Psychology*, *66*(12), 2426-2442.

Bonnet, L., Comte, A., Tatu, L., Millot, J.-L., Moulin, T., & Medeiros de Bustos, E. (2015). The role of the amygdala in the perception of positive emotions : an "intensity detector". *Frontiers in behavioral neuroscience*, *9*, 178.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185-207.

Boucart, M., Lenoble, Q., Quettelart, J., Szaffarczyk, S., Despretz, P., & Thorpe, S. J. (2016). Finding faces, animals, and vehicles in far peripheral vision. *Journal of Vision*, *16*(2), 10.

Boynton, G. M. (2005). Contrast gain in the brain. *Neuron*, *47*(4), 476-477.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, *16*(13), 4207-4221.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436.

Breitmeyer, B. G. (2014). Contributions of magno-and parvocellular channels to conscious and non-conscious vision. *Philosophical Transactions of the Royal Society B : Biological Sciences*, *369*(1641), 20130213.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99-108.

Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli : A review. *Cognition and emotion*, *24*(3), 377-400.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, *77*(3), 305-327.

Burra, N., Hervais-Adelman, A., Celeghin, A., De Gelder, B., & Pegna, A. J. (2019). Affective blindsight relies on low spatial frequencies. *Neuropsychologia*, *128*, 44-49.

Buswell, G. T. (1935). How people look at pictures : a study of the psychology and perception in art.

Bylinskii, Z., Recasens, A., Borji, A., Oliva, A., Torralba, A., & Durand, F. (2016). Where should saliency models look next ? *European Conference on Computer Vision*, 809-824.

Calder, A. J. (1996). Facial emotion recognition after bilateral amygdala damage : Differentially severe impairment of fear. *Cognitive Neuropsychology*, *13*(5), 699-745.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology : Human perception and performance*, *26*(2), 527.

Calvo, M. G., Fernández-Martın, A., & Nummenmaa, L. (2012). Perceptual, categorical, and affective processing of ambiguous smiling facial expressions. *Cognition*, *125*(3), 373-393.

Calvo, M. G., Gutiérrez-Garcıa, A., Fernández-Martın, A., & Nummenmaa, L. (2014). Recognition of facial expressions of emotion is related to their frequency in everyday life. *Journal of Nonverbal Behavior*, *38*(4), 549-567.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109-115.

Calvo, M. G., & Marrero, H. (2009). Visual search of emotional faces: the role of affective content and featural distinctiveness. *Cognition & Emotion*, *23*(4), 782-806.

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces : salient physical features guide effective visual search. *Journal of Experimental Psychology : General*, *137*(3), 471.

Calvo, M. G., & Nummenmaa, L. (2009). Eye-movement assessment of the time course in facial expression recognition: neurophysiological implications. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(4), 398-411.

Calvo, M. G., & Nummenmaa, L. (2011). Time course of discrimination between emotional facial expressions: the role of visual saliency. *Vision Research*, *51*(15), 1751-1759.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cognition and Emotion*, *30*(6), 1081-1106.

Calvo, M. G., Nummenmaa, L., & Avero, P. (2008). Visual search of emotional faces: eye-movement assessment of component processes. *Experimental Psychology*, *55*(6), 359-370.

Campagne, A., Fradcourt, B., Pichat, C., Baciu, M., Kauffmann, L., & Peyrin, C. (2016). Cerebral correlates of emotional and action appraisals during visual processing of emotional scenes depending on spatial frequency : A pilot study. *Plos One*, *11*(1), e0144393.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of physiology*, *197*(3), 551.

Cangöz, B., Altun, A., Aşkar, P., Baran, Z., & Mazman, S. G. (2013). Examining the visual screening patterns of emotional facial expressions with gender, age and lateralization. *Journal of Eye Movement Research*, *6*(4).

Carretié, L. (2014). Exogenous (automatic) attention to emotional stimuli: a review. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(4), 1228-1258.

Carretié, L., Hinojosa, J. A., López-Martın, S., Albert, J., Tapia, M., & Pozo, M. A. (2009). Danger is worse when it moves : Neural and behavioral indices of enhanced attentional capture by dynamic threatening stimuli. *Neuropsychologia*, *47*(2), 364-369.

Celeghin, A., Bagnis, A., Diano, M., Mendez, C. A., Costa, T., & Tamietto, M. (2019). Functional neuroanatomy of blindsight revealed by activation likelihood estimation meta-analysis. *Neuropsychologia*, *128*, 109-118.

Celeghin, A., de Gelder, B., & Tamietto, M. (2015). From affective blindsight to emotional consciousness. *Consciousness and cognition*, *36*, 414-425.

Celeghin, A., Diano, M., Bagnis, A., Viola, M., & Tamietto, M. (2017). Basic emotions in human neuroscience : neuroimaging and beyond. *Frontiers in Psychology*, *8*, 1432.

Cellerino, A., Borghetti, D., & Sartucci, F. (2004). Sex differences in face gender recognition in humans. *Brain Research Bulletin*, *63*(6), 443-449.

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: experimental data and computer model. *Journal of Vision*, *9*(12), 10-10.

Chen, C.-Y., Sonnenberg, L., Weller, S., Witschel, T., & Hafed, Z. M. (2018). Spatial frequency sensitivity in macaque midbrain. *Nature communications*, *9*(1), 1-13.

Chen, Y., Li, H., Jin, Z., Shou, T., & Yu, H. (2014). Feedback of the amygdala globally modulates visual response of primary visual cortex in the cat. *Neuroimage*, *84*, 775-785.

Coe, B. C., Trappenberg, T., & Munoz, D. P. (2019). Modeling saccadic action selection: cortical and basal ganglia signals coalesce in the superior colliculus. *Frontiers in Systems Neuroscience*, *13*, 3.

Coelho, C. M., Cloete, S., & Wallis, G. (2010). The face-in-the-crowd effect : When angry faces are just cross (es). *Journal of Vision*, *10*(1), 7-7.

Collins, T., & Doré-Mazars, K. (2006). Eye movement signals influence perception : evidence from the adaptation of reactive and volitional saccades. *Vision research*, *46*(21), 3659-3673.

Collins, T., Heed, T., & Röder, B. (2010). Visual target selection and motor planning define attentional enhancement at perceptual processing stages. *Frontiers in human neuroscience*, *4*, 14.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Petersen, S. E., Raichle, M. E., Van Essen, D. C., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761-773.

Corradi-Dell'Acqua, C., Schwartz, S., Meaux, E., Hubert, B., Vuilleumier, P., & Deruelle, C. (2014). Neural responses to emotional expression information in high-and low-spatial frequency in autism : evidence for a cortical dysfunction. *Frontiers in human neuroscience*, *8*, 189.

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 5.

Crouzet, S. M. (2010). Fast saccades toward faces: face detection in just 100 ms. *Journal of Vision*, *10*(4), 1-17.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces : Face detection in just 100 ms. *Journal of Vision*, *10*(4), 16-16.

Crouzet, S. M., & Thorpe, S. J. (2011). Low-Level Cues and Ultra-Fast Face Detection. *Frontiers in Psychology*, *2*.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: a neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience*, *14*(8), 1158-1173.

Daniel, P., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology*, *159*(2), 203-221.

Darwin, C. (1872). *The expression of the emotions in man and animals*.

Davies-Thompson, J., Newling, K., & Andrews, T. J. (2013). Image-invariant responses in face-selective regions do not explain the perceptual advantage for familiar face recognition. *Cerebral Cortex*, *23*(2), 370-377.

Davis, M., & Whalen, P. J. (2001). The amygdala: vigilance and emotion. *Molecular Psychiatry*, *6*(1), 13-34.

Day-Brown, J. D., Wei, H., Chomsung, R. D., Petry, H. M., & Bickford, M. E. (2010). Pulvinar projections to the striatum and amygdala in the tree shrew. *Frontiers in neuroanatomy*, *4*, 143.

De Franceschi, G., & Solomon, S. G. (2020). Dynamic contextual modulation in superior colliculus of awake mouse. *Eneuro*, *7*(5).

De Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. *Neuroreport*, *10*(18), 3759-3763.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision research*, *22*(5), 545-559.

De Valois, R. L., Morgan, H., & Snodderly, D. M. (1974). Psychophysical studies of monkey vision-III. Spatial luminance contrast sensitivity tests of macaque and human observers. *Vision research*, *14*(1), 75-81.

de Gelder, B., van Honk, J., & Tamietto, M. (2011). Emotion in the brain : of low roads, high roads and roads less travelled. *Nature Reviews Neuroscience*, *12*(7), 425-425.

de Haan, E. H., & Cowey, A. (2011). On the usefulness of 'what'and 'where'pathways in vision. *Trends in cognitive sciences*, *15*(10), 460-466.

de Haan, E. H., Jackson, S. R., Schenk, T., et al. (2018). Where are we now with'What'and'How'? *Cortex*, *98*.

de Lissa, P., Sokhn, N., Lasrado, S., Tanaka, K., Watanabe, K., & Caldara, R. (2021). Rapid saccadic categorization of other-race faces. *Journal of Vision*, *21*(12), 1-1.

Demos, K. E., Kelley, W. M., Ryan, S. L., Davis, F. C., & Whalen, P. (2008). Human amygdala sensitivity to the pupil size of others. *Cerebral cortex*, *18*(12), 2729-2734.

Devue, C., Belopolsky, A. V., & Theeuwes, J. (2012). Oculomotor guidance and capture by irrelevant faces. *PloS one*, *7*(4), e34598.

Devue, C., & Grimshaw, G. M. (2017). Faces are special, but facial expressions aren't: insights from an oculomotor capture paradigm. *Attention, Perception, & Psychophysics*, *79*(5), 1438-1452.

D'Hondt, F., Szaffarczyk, S., Sequeira, H., & Boucart, M. (2016). Explicit and implicit emotional processing in peripheral vision: a saccadic choice paradigm. *Biological Psychology*, *119*, 91-100.

Diano, M., Celeghin, A., Bagnis, A., & Tamietto, M. (2017). Amygdala response to emotional stimuli without awareness: facts and interpretations. *Frontiers in Psychology*, *7*.

Dobs, K., Bülthoff, I., & Schultz, J. (2018). Use and usefulness of dynamic face stimuli for face perception studies—A review of behavioral findings and methodology. *Frontiers in psychology*, 1355.

Dogra, A., & Bhalla, P. (2014). Image sharpening by gaussian and butterworth high pass filter. *Biomedical and Pharmacology Journal*, *7*(2), 707-713.

Dorris, M. C., Olivier, E., & Munoz, D. P. (2007). Competitive integration of visual and preparatory signals in the superior colliculus during saccadic programming. *Journal of Neuroscience*, *27*(19), 5053-5062.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual review of vision science*, *1*, 393-416.

Duffy, E. (1962). Activation and behavior.

Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, *38*(4), 659-671.

Edelman, J. A., & Keller, E. L. (1998). Dependence on target configuration of express saccade-related activity in the primate superior colliculus. *Journal of Neurophysiology*, *80*(3), 1407-1426.

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion*, *11*(4), 860-865.

Eisenbarth, H., Alpers, G. W., Segrè, D., Calogero, A., & Angrilli, A. (2008). Categorization and evaluation of emotional faces in psychopathic women. *Psychiatry research*, *159*(1-2), 189-195.

Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion*, *98*(45-60), 16.

Ekman, P. (2002). Facial action coding system (FACS). *A human face*.

Ekman, P. (2004). Emotions revealed. *Bmj*, *328*(Suppl S5).

Ekman, P. (2006). *Darwin and facial expression : A century of research in review*. Ishk.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion review*, *3*(4), 364-370.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124.

Ekman, P., & Friesen, W. V. (1978). Manual of the facial action coding system (FACS). *Trans. ed. Vol. Consulting Psychologists Press, Palo Alto*.

Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy : cultural exposure and facial emotion recognition. *Journal of personality and social psychology*, *85*(2), 276.

Emery, N. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581-604.

Entzmann, L., Guyader, N., Kauffmann, L., Lenouvel, J., Charles, C., Peyrin, C., Vuillaume, R., & Mermillod, M. (2021). The Role of Emotional Content and Perceptual Saliency During the Programming of Saccades Toward Faces. *Cognitive Science*, *45*(10), e13042.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., et al. (2016). The human brainnetome atlas : a new brain atlas based on connectional architecture. *Cerebral cortex*, *26*(8), 3508-3526.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect ? *Journal of Experimental Psychology : Human perception and performance*, *21*(3), 628.

Fecteau, J., & Munoz, D. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382-390.

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of personality and social psychology*, *74*(4), 967.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), 2379.

Findlay, J. M. (1982). Global visual processing for saccadic eye movements. *Vision Research*, *22*(8), 1033-1045.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661-674.

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions ? *Journal of Vision*, *11*(3), 17-17.

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, *16*(3), 553-567.

Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat: amygdala reactivity across multiple expressions of facial affect. *NeuroImage*, *30*(4), 1441-1448.

Flevaris, A. V., Bentin, S., & Robertson, L. C. (2011). Attention to hierarchical level influences attentional selection of spatial scale. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 12-22.

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological science*, *18*(12), 1050-1057.

Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. (2009). The correlates of subjective perception of identity and expression in the face network : an fMRI adaptation study. *Neuroimage*, *44*(2), 569-580.

Fox, E., Lester, V., Russo, R., Bowles, R., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion : Are angry faces detected more efficiently ? *Cognition & emotion*, *14*(1), 61-92.

Frank, D., Dewitt, M., Hudgens-Haney, M., Schaeffer, D., Ball, B., Schwarz, N., Hussein, A., Smart, L., & Sabatinelli, D. (2014). Emotion regulation : quantitative meta-analysis of functional activation and deactivation. *Neuroscience & Biobehavioral Reviews*, *45*, 202-211.

Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological Bulletin*, *134*(5), 662-676.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging : a general linear approach. *Human brain mapping*, *2*(4), 189-210.

Furl, N., Henson, R. N., Friston, K. J., & Calder, A. J. (2013). Top-down control of visual responses to fear by the amygdala. *Journal of Neuroscience*, *33*(44), 17435-17443.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., Benedetti, F., Abbamonte, M., Gasparotti, R., Barale, F., et al. (2009). Functional atlas of emotional faces processing : a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of psychiatry & neuroscience*.

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., & Goodale, M. A. (2005). The involvement of the "fusiform face area" in processing facial expression. *Neuropsychologia*, *43*(11), 1645-1654.

Ganis, G., Smith, D., & Schendan, H. E. (2012). The N170, not the P1, indexes the earliest time for categorical perception of faces, regardless of interstimulus variance. *Neuroimage*, *62*(3), 1563-1574.

Garavan, H., Pendergrass, J. C., Ross, T. J., Stein, E. A., & Risinger, R. C. (2001). Amygdala response to both positively and negatively valenced stimuli. *Neuroreport*, *12*(12), 2779-2783.

Garvert, M. M., Friston, K. J., Dolan, R. J., & Garrido, M. I. (2014). Subcortical amygdala pathways enable rapid face processing. *NeuroImage*, *102*, 309-316.

George, N., Driver, J., & Dolan, R. J. (2001). Seen gaze-direction modulates fusiform activity and its coupling with other brain areas during face processing. *Neuroimage*, *13*(6), 1102-1112.

Ghodrati, M., Khaligh-Razavi, S.-M., & Lehky, S. R. (2017). Towards building a more complex view of the lateral geniculate nucleus : recent advances in understanding its role. *Progress in Neurobiology*, *156*, 214-255.

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska directed emotional faces : a validation study. *Cognition and emotion*, *22*(6), 1094-1118.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, *34*(1), 77-86.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467-476.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"–holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 1023-1039.

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., Goode, R. A., Lee, D., Simmons, M., & Srinath, A. (2013). The efficiency of dynamic and static facial expression recognition. *Journal of Vision*, *13*(5), 23-23.

Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychological science*, *23*(4), 427-434.

Goldin, P. R., McRae, K., Ramel, W., & Gross, J. J. (2008). The neural bases of emotion regulation : reappraisal and suppression of negative emotion. *Biological psychiatry*, *63*(6), 577-586.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20-25.

Goodyear, B. G., & Menon, R. S. (1998). Effect of luminance contrast on BOLD fMRI response in human primary visual areas. *Journal of Neurophysiology*, *79*(4), 2204-2207.

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

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.*, *27*, 649-677.

Gross, D. M., & Preston, S. D. (2020). Darwin and the situation of emotion research. *Emotion Review*, *12*(3), 179-190.

Grühn, D., & Sharifian, N. (2016). Lists of emotional stimuli. In *Emotion measurement* (p. 145-164). Elsevier.

Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral cortex*, *22*(7), 1564-1576.

Guyader, N., Chauvin, A., Boucart, M., & Peyrin, C. (2017). Do low spatial frequencies explain the extremely fast saccades towards human faces? *Vision Research*, *133*, 100-111.

Habel, U., Windischberger, C., Derntl, B., Robinson, S., Kryspin-Exner, I., Gur, R. C., & Moser, E. (2007). Amygdala activation and facial expressions: explicit emotion discrimination versus implicit emotion processing. *Neuropsychologia*, *45*(10), 2369-2377.

Haxby, J. V., & Gobbini, M. I. (2011). *Distributed neural systems for face perception.* The Oxford Handbook of Face Perception.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223-233.

Hegde, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405-439.

Hernandez, N., Metzger, A., Magné, R., Bonnet-Brilhault, F., Roux, S., Barthelemy, C., & Martineau, J. (2009). Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia*, *47*(4), 1004-1012.

Hershler, O., & Hochstein, S. (2005). At first sight : A high-level pop out effect for faces. *Vision research*, *45*(13), 1707-1724.

Hershler, O., & Hochstein, S. (2006). With a careful look : Still no low-level confound to face pop-out. *Vision research*, *46*(18), 3028-3035.

Hess, U., Adams, R. B., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: are angry women more like men? *Journal of Vision*, *9*(12), 19-19.

Hess, U., & Thibault, P. (2009). Darwin and emotion expression. *American Psychologist*, *64*(2), 120.

Hinojosa, J., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression : A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *55*, 498-509.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature neuroscience*, *3*(1), 80-84.

Holland, P. C., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in cognitive sciences*, *3*(2), 65-73.

Holmes, A., Green, S., & Vuilleumier, P. (2005). The involvement of distinct visual channels in rapid attention towards fearful facial expressions. *Cognition & Emotion*, *19*(6), 899-922.

Holmes, A., Winston, J. S., & Eimer, M. (2005). The role of spatial frequency information for ERP components sensitive to faces and emotional facial expression. *Cognitive Brain Research*, *25*(2), 508-520.

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: fourier power spectrum biases ultrarapid face detection. *Journal of Vision*, *8*(12), 9-9.

Horstmann, G., Lipp, O. V., & Becker, S. I. (2012). Of toothy grins and angry snarls–open mouth displays contribute to efficiency gains in search for emotional faces. *Journal of Vision*, *12*(5), 7-7.

Horstmann, G., & Ansorge, U. (2009). Visual search for facial expressions of emotions : A comparison of dynamic and static faces. *Emotion*, *9*(1), 29.

Horstmann, G., & Becker, S. I. (2020). More efficient visual search for happy faces may not indicate guidance, but rather faster distractor rejection : Evidence from eye movements and fixations. *Emotion*, *20*(2), 206.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106-154.

Hubel, D. H., & Wiesel, T. N. (2004). *Brain and visual perception : the story of a 25-year collaboration*. Oxford University Press.

Huijgen, J., Dinkelacker, V., Lachat, F., Yahia-Cherif, L., El Karoui, I., Lemaréchal, J.-D., Adam, C., Hugueville, L., & George, N. (2015). Amygdala processing of social cues from faces : an intracrebral EEG study. *Social Cognitive and Affective Neuroscience*, *10*(11), 1568-1576.

Hunt, A. R., Cooper, R. M., Hungr, C., & Kingstone, A. (2007). The effect of emotional faces on eye movements and attention. *Visual Cognition*, *15*(5), 513-531.

Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision research*, *51*(10), 1192-1205.

Inagaki, T. K., Muscatell, K. A., Irwin, M. R., Cole, S. W., & Eisenberger, N. I. (2012). Inflammation selectively enhances amygdala activity to socially threatening images. *Neuroimage*, *59*(4), 3222-3226.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254-1259.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current biology*, *19*(18), 1543-1548.

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, *109*(19), 7241-7244.

Janak, P. H., & Tye, K. M. (2015). From circuits to behaviour in the amygdala. *Nature*, *517*(7534), 284-292.

Janik, S. W., Wellens, A. R., Goldberg, M. L., & Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces. *Perceptual and motor skills*, *47*(3), 857-858.

Jeantet, C., Caharel, S., Schwan, R., Lighezzolo-Alnot, J., & Laprevote, V. (2018). Factors influencing spatial frequency extraction in faces: a review. *Neuroscience & Biobehavioral Reviews*, *93*, 123-138.

Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, *78*(1), 193-202.

Jessen, S., & Grossmann, T. (2017). Exploring the role of spatial frequency information during neural emotion processing in human infants. *Frontiers in Human Neuroscience*, *11*, 486.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, *6*(10), 766-774.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1-19.

Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, *50*, 169-179.

Jones, E., & Burton, H. (1976). A projection from the medial pulvinar to the amygdala in primates. *Brain research*, *104*(1), 142-147.

Joukal, M. (2017). Anatomy of the human visual pathway. In *Homonymous visual field defects* (p. 1-16). Springer.

Juth, P., Lundqvist, D., Karlsson, A., & Öhman, A. (2005). Looking for foes and friends : perceptual and emotional factors when finding a face in the crowd. *Emotion*, *5*(4), 379.

Kadosh, K. C., Walsh, V., & Kadosh, R. C. (2011). Investigating face-property specific processing in the right OFA. *Social cognitive and affective neuroscience*, *6*(1), 58-65.

Kanat, M., Heinrichs, M., Mader, I., Van Elst, L. T., & Domes, G. (2015). Oxytocin modulates amygdala reactivity to masked fearful eyes. *Neuropsychopharmacology*, *40*(11), 2632-2638.

Kaplan, E. (2004). The M, P, and K pathways of the primate visual system. *The visual neurosciences*, *1*, 481-493.

Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, *107*, 49-57.

Kauffmann, L., Khazaz, S., Peyrin, C., & Guyader, N. (2021). Isolated face features are sufficient to elicit ultra-rapid and involuntary orienting responses toward faces. *Journal of Vision*, *21*(2), 4.

Kauffmann, L., Peyrin, C., Chauvin, A., Entzmann, L., Breuil, C., & Guyader, N. (2019). Face perception influences the programming of eye movements. *Scientific Reports*, *9*(1).

Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, *112*, 86-95.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in integrative neuroscience*, *8*, 37.

Kawasaki, H., Tsuchiya, N., Kovach, C. K., Nourski, K. V., Oya, H., Howard, M. A., & Adolphs, R. (2012). Processing of facial emotion in the human fusiform gyrus. *Journal of cognitive neuroscience*, *24*(6), 1358-1370.

Kennedy, D. P., & Adolphs, R. (2010). Impaired fixation to eyes following amygdala damage arises from abnormal bottom-up attention. *Neuropsychologia*, *48*(12), 3392-3398.

Kennerley, S. W., Behrens, T. E., & Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature neuroscience*, *14*(12), 1581-1589.

Kim, M. J., Solomon, K. M., Neta, M., Davis, F. C., Oler, J. A., Mazzulla, E. C., & Whalen, P. J. (2016). A face versus non-face context influences amygdala responses to masked fearful eye whites. *Social cognitive and affective neuroscience*, *11*(12), 1933-1941.

Kim, U. S., Mahroo, O. A., Mollon, J. D., & Yu-Wai-Man, P. (2021). Retinal ganglion cells—diversity of cell types and clinical relevance. *Frontiers in Neurology*, 635.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research*, *46*(11), 1762-1776.

Klink, P. C., Jentgens, P., & Lorteije, J. A. M. (2014). Priority maps explain the roles of value, attention, and salience in goal-oriented behavior. *Journal of Neuroscience*, *34*(42), 13867-13869.

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*(6635), 767-768.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2010). Facial emotion perception in schizophrenia : a meta-analytic review. *Schizophrenia bulletin*, *36*(5), 1009-1019.

Koller, K., Rafal, R. D., Platt, A., & Mitchell, N. D. (2019). Orienting toward threat : Contributions of a subcortical pathway transmitting retinal afferents to the amygdala via the superior colliculus and pulvinar. *Neuropsychologia*, *128*, 78-86.

Koster, E. H., Verschuere, B., Burssens, B., Custers, R., & Crombez, G. (2007). Attention for emotional faces under restricted awareness revisited : Do emotional faces automatically attract attention ? *Emotion*, *7*(2), 285.

Krasovskaya, S., & MacInnes, W. J. (2019). Salience models : A computational cognitive neuroscience review. *Vision*, *3*(4), 56.

Krause, F. C., Linardatos, E., Fresco, D. M., & Moore, M. T. (2021). Facial emotion recognition in major depressive disorder : A meta-analytic review. *Journal of Affective Disorders*, *293*, 320-328.

Krolak-Salmon, P., Hénaff, M.-A., Vighetto, A., Bertrand, O., & Mauguière, F. (2004). Early amygdala reaction to fear spreading in occipital, temporal, and frontal cortex : a depth electrode ERP study in human. *Neuron*, *42*(4), 665-676.

Krumhuber, E. G., Küster, D., Namba, S., Shah, D., & Calvo, M. G. (2021). Emotion recognition from posed and spontaneous dynamic expressions : Human observers versus machine analysis. *Emotion*, *21*(2), 447.

Kulke, L. (2019). Neural mechanisms of overt attention shifts to emotional faces. *Neuroscience*, *418*, 59-68.

Kumar, D., & Srinivasan, N. (2011). Emotion perception is mediated by spatial frequency content. *Emotion*, *11*(5), 1144-1151.

Lacruz, R. S., Stringer, C. B., Kimbel, W. H., Wood, B., Harvati, K., O'Higgins, P., Bromage, T. G., & Arsuaga, J.-L. (2019). The evolutionary history of the human face. *Nature Ecology & Evolution*, *3*(5), 726-736.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and emotion*, *24*(8), 1377-1388.

LeDoux, J. (1998). *The emotional brain : The mysterious underpinnings of emotional life*. Simon ; Schuster.

LeDoux, J., Cicchetti, P., Xagoraris, A., & Romanski, L. M. (1990). The lateral amygdaloid nucleus : sensory interface of the amygdala in fear conditioning. *Journal of neuroscience*, *10*(4), 1062-1069.

LeDoux, J., Sakaguchi, A., & Reis, D. J. (1984). Subcortical efferent projections of the medial geniculate nucleus mediate emotional responses conditioned to acoustic stimuli. *Journal of Neuroscience*, *4*(3), 683-698.

Leopold, D. A., & Rhodes, G. (2010). A comparative view of face perception. *Journal of Comparative Psychology*, *124*(3), 233-251.

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological Research Psychologische Forschung*, *69*(1), 22-29.

Lesmes, L. A., Lu, Z.-L., Baek, J., & Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function : The quick CSF method. *Journal of vision*, *10*(3), 17-17.

Lewis, M. B., & Edmonds, A. J. (2003). Face detection : Mapping human performance. *Perception*, *32*(8), 903-920.

Li, R., & Cottrell, G. (2012). A new angle on the EMPATH model : Spatial frequency orientation in recognition of facial expressions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*(34).

Liddell, B. J., Brown, K. J., Kemp, A. H., Barton, M. J., Das, P., Peduto, A., Gordon, E., & Williams, L. M. (2005). A direct brainstem–amygdala–cortical 'alarm'system for subliminal signals of fear. *Neuroimage*, *24*(1), 235-243.

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research : re-balancing the scale. *Social cognitive and affective neuroscience*, *4*(4), 423-428.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion : a meta-analytic review. *The Behavioral and brain sciences*, *35*(3), 121.

Lipp, O. V., Price, S. M., & Tellegen, C. L. (2009). No effect of inversion on attentional and affective processing of facial expressions. *Emotion*, *9*(2), 248-259.

Liu, M., Liu, C. H., Zheng, S., Zhao, K., & Fu, X. (2021). Reexamining the neural network involved in perception of facial expression : a meta-analysis. *Neuroscience & Biobehavioral Reviews*, *131*, 179-191.

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away ? *Psychonomic Bulletin & Review*, *12*(1), 43-65.

Lucas, N., & Vuilleumier, P. (2008). Effects of emotional and non-emotional cues on visual search in neglect patients : Evidence for distinct sources of attentional guidance. *Neuropsychologia*, *46*(5), 1401-1414.

Lundqvist, D., Bruce, N., & Öhman, A. (2015). Finding an emotional face in a crowd : Emotional and perceptual stimulus factors influence visual search efficiency. *Cognition and Emotion*, *29*(4), 621-633.

Lundqvist, D., Flykt, A., & Ohman, A. (1998). The Karolinska directed emotional faces (KDEF).

Lundqvist, D., Juth, P., & Öhman, A. (2014). Using facial emotional stimuli in visual search experiments : The arousal factor explains contradictory results. *Cognition and Emotion*, *28*(6), 1012-1029.

Lupp, U., Hauske, G., & Wolf, W. (1976). Perceptual latencies to sinusoidal gratings. *Vision research*, *16*(9), 969-972.

Maior, R. S., Hori, E., Tomaz, C., Ono, T., & Nishijo, H. (2010). The monkey pulvinar neurons differentially respond to emotional expressions of human faces. *Behavioural brain research*, *215*(1), 129-135.

Makandar, A., & Halalli, B. (2015). Image enhancement techniques using highpass and lowpass filters. *International Journal of Computer Applications*, *109*(14), 12-15.

Marat, S., Rahman, A., Pellerin, D., Guyader, N., & Houzet, D. (2013). Improving visual saliency by adding 'face feature map'and 'center bias'. *Cognitive Computation*, *5*(1), 63-75.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature reviews neuroscience*, *5*(3), 229-240.

Masland, R. H. (2001). The fundamental plan of the retina. *Nature neuroscience*, *4*(9), 877-886.

Masland, R. H. (2012). The neuronal organization of the retina. *Neuron*, *76*(2), 266-280.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences*, *6*(6), 255-260.

McFadyen, J., Mattingley, J. B., & Garrido, M. I. (2019). An afferent white matter pathway from the pulvinar to the amygdala facilitates fear recognition. *eLife*, *8*, e40766.

McFadyen, J., Mermillod, M., Mattingley, J. B., Halász, V., & Garrido, M. I. (2017). A rapid subcortical amygdala route for faces irrespective of spatial frequency and emotion. *The Journal of Neuroscience*, *37*(14), 3864-3874.

McIntosh, R. D., & Schenk, T. (2009). Two visual streams for perception and action : current trends. *Neuropsychologia*, *47*(6), 1391-1396.

McLellan, T., Johnston, L., Dalrymple-Alford, J., & Porter, R. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cognition & Emotion*, *24*(8), 1277-1292.

McPeek, R. M., Skavenski, A. A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision Research*, *40*(18), 2499-2516.

Meletti, S., Cantalupo, G., Benuzzi, F., Mai, R., Tassi, L., Gasparini, E., Tassinari, C. A., & Nichelli, P. (2012). Fear and happiness in the eyes : An intra-cerebral event-related potential study from the human amygdala. *Neuropsychologia*, *50*(1), 44-54.

Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y. H., Vuilleumier, P., Gil-Nagel, A., & Strange, B. A. (2016). A fast pathway for fear in human amygdala. *Nature Neuroscience*, *19*(8), 1041-1049.

Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions: evidence from neural computation. *Neurocomputing*, *73*(13), 2522-2531.

Mermillod, M., Bourrier, Y., David, E., Kauffmann, L., Chauvin, A., Guyader, N., Dutheil, F., & Peyrin, C. (2019). The importance of recurrent top-down synaptic connections for the anticipation of dynamic emotions. *Neural Networks*, *109*, 19-30.

Mermillod, M., Vermeulen, N., Lundqvist, D., & Niedenthal, P. M. (2009). Neural computation as a tool to differentiate perceptual from emotional processes: the case of anger superiority effect. *Cognition*, *110*(3), 346-357.

Mimeault, D., Paquet, V., Molotchnikoff, S., Lepore, F., & Guillemot, J.-P. (2004). Disparity sensitivity in the superior colliculus of the cat. *Brain research*, *1010*(1-2), 87-94.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research*, *6*(1), 57-77.

Moors, A., & De Houwer, J. (2006). Automaticity : a theoretical and conceptual analysis. *Psychological bulletin*, *132*(2), 297.

Morris, J. S., deBonis, M., & Dolan, R. J. (2002). Human amygdala responses to fearful eyes. *Neuroimage*, *17*(1), 214-222.

Morris, J. S., Friston, K. J., Büchel, C., Frith, C. D., Young, A. W., Calder, A. J., & Dolan, R. J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain : a journal of neurology*, *121*(1), 47-57.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, *383*(6603), 812-815.

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences*, *96*(4), 1680-1685.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic bulletin & review*, *8*(3), 454-469.

Mothes-Lasch, M., Mentzel, H.-J., Miltner, W. H., & Straube, T. (2013). Amygdala activation to fearful faces under attentional load. *Behavioural brain research*, *237*, 172-175.

Mulckhuyse, M. (2018). The influence of emotional stimuli on the oculomotor system: a review of the literature. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(3), 411-425.

Mulckhuyse, M., Crombez, G., & Van der Stigchel, S. (2013). Conditioned fear modulates visual selection. *Emotion*, *13*(3), 529.

Murphy, J., Gray, K. L., & Cook, R. (2017). The composite face illusion. *Psychonomic Bulletin & Review*, *24*(2), 245-261.

Musel, B., Bordier, C., Dojat, M., Pichat, C., Chokron, S., Le Bas, J.-F., & Peyrin, C. (2013). Retinotopic and lateralized processing of spatial frequencies in human visual cortex during scene categorization. *Journal of Cognitive Neuroscience*, *25*(8), 1315-1331.

Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? (C. Alain, Éd.). *PLoS ONE*, *7*(6), e38493.

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of cognitive neuroscience*, *26*(10), 2287-2297.

Nakashima, T., Kaneko, K., Goto, Y., Abe, T., Mitsudo, T., Ogata, K., Makinouchi, A., & Tobimatsu, S. (2008). Early ERP components differentially extract facial features: evidence for spatial frequency-and-contrast detectors. *Neuroscience Research*, *62*(4), 225-235.

Navon, D. (1977). Forest before trees : The precedence of global features in visual perception. *Cognitive psychology*, *9*(3), 353-383.

Nguyen, M. N., Hori, E., Matsumoto, J., Tran, A. H., Ono, T., & Nishijo, H. (2013). Neuronal responses to face-like stimuli in the monkey pulvinar. *European Journal of Neuroscience*, *37*(1), 35-51.

Nguyen, M. N., Matsumoto, J., Hori, E., Maior, R. S., Tomaz, C., Tran, A. H., Ono, T., & Nishijo, H. (2014). Neuronal responses to face-like and facial stimuli in the monkey superior colliculus. *Frontiers in Behavioral Neuroscience*, *8*.

Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In *Extrastriate cortex in primates* (p. 205-241). Springer.

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, *6*(2), 257-268.

Öhman, A., Juth, P., & Lundqvist, D. (2010). Finding the face in a crowd : Relationships between distractor redundancy, target emotion, and target gender. *Cognition and Emotion*, *24*(7), 1216-1228.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*(3), 381-396.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483-522.

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.

Ottaviani, C., Cevolani, D., Nucifora, V., Borlimi, R., Agati, R., Leonardi, M., De Plato, G., & Brighetti, G. (2012). Amygdala responses to masked and low spatial frequency fearful faces : a preliminary fMRI study in panic disorder. *Psychiatry Research : Neuroimaging*, *203*(2-3), 159-165.

Ouellette, B. G., & Casanova, C. (2006). Overlapping visual response latency distributions in visual cortices and LP-pulvinar complex of the cat. *Experimental brain research*, *175*(2), 332-341.

Ozgen, E., Payne, H. E., Sowden, P. T., & Schyns, P. G. (2006). Retinotopic sensitisation to spatial scale : evidence for flexible spatial frequency processing in scene perception. *Vision Research*, *46*(6-7), 1108-1119.

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression : Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 634-638.

Palermo, R., & Rhodes, G. (2007). Are you always on my mind ? A review of how face perception and attention interact. *Neuropsychologia*, *45*(1), 75-92.

Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision research*, *90*, 10-14.

Perfetto, S., Wilder, J., & Walther, D. B. (2020). Effects of spatial frequency filtering choices on the perception of filtered images. *Vision*, *4*(2), 29.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature reviews neuroscience*, *9*(2), 148-158.

Pessoa, L. (2010). Emotion and cognition and the amygdala : from "what is it ?" to "what's to be done ?" *Neuropsychologia*, *48*(12), 3416-3429.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience*, *11*(11), 773-782.

Peters, J. C., Goebel, R., & Goffaux, V. (2018). From coarse to fine : Interactive feature processing precedes local feature analysis in human face perception. *Biological psychology*, *138*, 1-10.

Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, *186*, 103-112.

Peyrin, C., Baciu, M., Segebarth, C., & Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *Neuroimage*, *23*(2), 698-707.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion : a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, *16*(2), 331-348.

Phelps, E. A. (2004). Human emotion and memory : interactions of the amygdala and hippocampal complex. *Current opinion in neurobiology*, *14*(2), 198-202.

Phelps, E. A. (2006). Emotion and cognition : insights from studies of the human amygdala. *Annu. Rev. Psychol.*, *57*, 27-53.

Phillips, M. L., Young, A. W., Scott, S., Calder, A. J., Andrew, C., Giampietro, V., Williams, S. C., Bullmore, E. T., Brammer, M., & Gray, J. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society of London. Series B : Biological Sciences*, *265*(1408), 1809-1817.

Piech, R. M., McHugo, M., Smith, S. D., Dukic, M. S., Van Der Meer, J., Abou-Khalil, B., Most, S. B., & Zald, D. H. (2011). Attentional capture by emotional stimuli is preserved in patients with amygdala lesions. *Neuropsychologia*, *49*(12), 3314-3319.

Piech, R. M., McHugo, M., Smith, S. D., Dukic, M. S., Van Der Meer, J., Abou-Khalil, B., & Zald, D. H. (2010). Fear-enhanced visual search persists after amygdala lesions. *Neuropsychologia*, *48*(12), 3430-3435.

Piepers, D., & Robbins, R. (2012). A review and clarification of the terms "holistic,""configural," and "relational" in the face perception literature. *Frontiers in psychology*, *3*, 559.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage*, *56*(4), 2356-2363.

Pöppel, E., Held, R., & Frost, D. (1973). Residual visual function after brain wounds involving the central visual pathways in man. *Nature*, *243*(5405), 295-296.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3-25.

Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., & Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: time course and topographic evoked-potentials mapping. *Human Brain Mapping*, *26*(1), 65-79.

Pourtois, G., Spinelli, L., Seeck, M., & Vuilleumier, P. (2010). Temporal precedence of emotion over attention modulations in the lateral amygdala : Intracranial ERP evidence from a patient with temporal lobe epilepsy. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(1), 83-93.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of neuroscience*, *18*(6), 2188-2199.

Puls, S., & Rothermund, K. (2018). Attending to emotional expressions : no evidence for automatic capture in the dot-probe task. *Cognition and Emotion*, *32*(3), 450-463.

Quek, G. L., Liu-Shuang, J., Goffaux, V., & Rossion, B. (2018). Ultra-coarse, single-glance human face detection in a dynamic visual stream. *NeuroImage*, *176*, 465-476.

Rafal, R. D., Koller, K., Bultitude, J. H., Mullins, P., Ward, R., Mitchell, A. S., & Bell, A. H. (2015). Connectivity between the superior colliculus and the amygdala in humans and macaque monkeys : virtual dissection with probabilistic DTI tractography. *Journal of neurophysiology*, *114*(3), 1947-1962.

Recio, G., Schacht, A., & Sommer, W. (2013). Classification of dynamic facial expressions of emotion presented briefly. *Cognition & emotion*, *27*(8), 1486-1494.

Reynolds, M. G., Eastwood, J. D., Partanen, M., Frischen, A., & Smilek, D. (2009). Monitoring eye movements while searching for affective faces. *Visual Cognition*, *17*(3), 318-333.

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological bulletin*, *140*(5), 1281.

Rigoulot, S., D'Hondt, F., Defoort-Dhellemmes, S., Despretz, P., Honoré, J., & Sequeira, H. (2011). Fearful faces impact in peripheral vision : behavioral and neural evidence. *Neuropsychologia*, *49*(7), 2013-2021.

Rigoulot, S., D'Hondt, F., Honore, J., & Sequeira, H. (2012). Implicit emotional processing in peripheral vision : Behavioral and neural evidence. *Neuropsychologia*, *50*(12), 2887-2896.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians : evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1), 31-40.

Robertson, L. C., & Ivry, R. (2000). Hemispheric asymmetries: attention to visual and auditory primitives. *Current Directions in Psychological Science*, *9*(2), 59-63.

Robinson, J. L., Laird, A. R., Glahn, D. C., Lovallo, W. R., & Fox, P. T. (2010). Metaanalytic connectivity modeling : delineating the functional connectivity of the human amygdala. *Human brain mapping*, *31*(2), 173-184.

Rodieck, R., & Watanabe, M. (1993). Survey of the morphology of macaque retinal ganglion cells that project to the pretectum, superior colliculus, and parvicellular laminae of the lateral geniculate nucleus. *Journal of Comparative Neurology*, *338*(2), 289-303.

Rolfs, M. (2015). Attention in active vision : A perspective on perceptual continuity across saccades. *Perception*, *44*(8-9), 900-919.

Rosenberg, E. L., & Ekman, P. (2020). *What the face reveals : Basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. Oxford University Press.

Rossion, B. (2009). Distinguishing the cause and consequence of face inversion : The perceptual field hypothesis. *Acta psychologica*, *132*(3), 300-312.

Rossion, B. (2013). The composite face illusion : A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*(2), 139-253.

Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends in cognitive sciences*, *18*(6), 310-318.

Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: disentangling the contribution of low-level visual cues from face perception. *Vision Research*, *51*(12), 1297-1311.

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? ten lessons on the n170. *NeuroImage*, *39*(4), 1959-1979.

Rossion, B., & Lochy, A. (2021). Is human face recognition lateralized to the right hemisphere due to neural competition with left-lateralized visual word recognition ? A critical review. *Brain Structure and Function*, 1-31.

Rossion, B., & Retter, T. L. (2020). Face Perception. *The cognitive neurosciences, Sixth Edition*.

Rucci, M., & Poletti, M. (2015). Control and functions of fixational eye movements. *Annual review of vision science*, *1*, 499.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression ? A review of the cross-cultural studies. *Psychological bulletin*, *115*(1), 102.

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., Beck, S., & Jeffries, J. (2011). Emotional perception : meta-analyses of face and natural scene processing. *Neuroimage*, *54*(3), 2524-2533.

Sabatinelli, D., Lang, P. J., Bradley, M. M., Costa, V. D., & Keil, A. (2009). The timing of emotional discrimination in human amygdala and ventral visual cortex. *Journal of Neuroscience*, *29*(47), 14864-14868.

Sadeh, B., Podlipsky, I., Zhdanov, A., & Yovel, G. (2010). Event-related potential and functional MRI measures of face-selectivity are highly correlated : a simultaneous ERP-fMRI investigation. *Human brain mapping*, *31*(10), 1490-1501.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala : an evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4), 303-316.

Sanders, M., Warrington, E., Marshall, J., & Wieskrantz, L. (1974). " Blindsight" : vision in a field defect. *The Lancet*, *303*(7860), 707-708.

Sassi, F., Campoy, G., Castillo, A., Inuggi, A., & Fuentes, L. J. (2014). Task difficulty and response complexity modulate affective priming by emotional facial expressions. *Quarterly Journal of Experimental Psychology*, *67*(5), 861-871.

Sato, S., & Kawahara, J. I. (2015). Attentional capture by completely task-irrelevant faces. *Psychological research*, *79*(4), 523-533.

Sato, W., Kochiyama, T., Uono, S., Matsuda, K., Usui, K., Inoue, Y., & Toichi, M. (2011). Rapid amygdala gamma oscillations in response to fearful facial expressions. *Neuropsychologia*, *49*(4), 612-617.

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion : an fMRI study. *Cognitive Brain Research*, *20*(1), 81-91.

Savage, R. A., Lipp, O. V., Craig, B. M., Becker, S. I., & Horstmann, G. (2013). In search of the emotional face : Anger versus happiness superiority in visual search. *Emotion*, *13*(4), 758.

Scheller, E., Büchel, C., & Gamer, M. (2012). Diagnostic features of emotional expressions are processed preferentially. *PloS one*, *7*(7), e41792.

Schiller, P. H., & Kendall, J. (2004). Temporal factors in target selection with saccadic eye movements. *Experimental Brain Research*, *154*(2), 154-159.

Schiller, P. H., & Malpeli, J. G. (1977). Properties and tectal projections of monkey retinal ganglion cells. *Journal of Neurophysiology*, *40*(2), 428-445.

Schindler, S., & Bublatzky, F. (2020). Attention and emotion : An integrative review of emotional face processing as a function of attention. *Cortex*.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of neurophysiology*, *79*(6), 3272-3278.

Schneider, K. A., & Kastner, S. (2005). Visual responses of the human superior colliculus : a high-resolution functional magnetic resonance imaging study. *Journal of neurophysiology*, *94*(4), 2491-2503.

Schubö, A., Gendolla, G. H. E., Meinecke, C., & Abele, A. E. (2006). Detecting emotional faces and features in a visual search paradigm: are faces special? *Emotion*, *6*(2), 246-256.

Schurgin, M., Nelson, J., Iida, S., Ohira, H., Chiao, J., & Franconeri, S. (2014). Eye movements during emotion recognition in faces. *Journal of vision*, *14*(13), 14-14.

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., & 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., Petro, L. S., & Smith, M. L. (2009). Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain : behavioral and brain evidence. *Plos one*, *4*(5), e5625.

Seymour, B., & Dolan, R. (2008). Emotion, decision making, and the amygdala. *Neuron*, *58*(5), 662-671.

Šimić, G., Tkalčić, M., Vukić, V., Mulc, D., Španić, E., Šagud, M., Olucha-Bordonau, F. E., Vukšić, M., & R Hof, P. (2021). Understanding emotions : Origins and roles of the amygdala. *Biomolecules*, *11*(6), 823.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Deep inside convolutional networks : Visualising image classification models and saliency maps. *Workshop at International Conference on Learning Representations*.

Smith, F. W., & Rossit, S. (2018). Identifying and detecting facial expressions of emotion in peripheral vision. *PloS one*, *13*(5), e0197160.

Smith, F. W., & Schyns, P. G. (2009). Smile through your fear and sadness: transmitting and identifying facial expression signals over a range of viewing distances. *Psychological Science*, *20*(10), 1202-1208.

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., & Merlusca, C. (2014). How task shapes the use of information during facial expression categorizations. *Emotion*, *14*(3), 478-487.

Snowden, R., Snowden, R. J., Thompson, P., & Troscianko, T. (2012). *Basic vision : an introduction to visual perception*. Oxford University Press.

Soares, S. C., Maior, R. S., Isbell, L. A., Tomaz, C., & Nishijo, H. (2017). Fast detector/first responder: interactions between the superior colliculus-pulvinar pathway and stimuli relevant to primates. *Frontiers in Neuroscience*, *11*.

Somerville, L. H., & Whalen, P. J. (2006). Prior experience as a stimulus category confound : an example using facial expressions of emotion. *Social cognitive and affective neuroscience*, *1*(3), 271-274.

Song, J., Liu, M., Yao, S., Yan, Y., Ding, H., Yan, T., Zhao, L., & Xu, G. (2017). Classification of Emotional Expressions Is Affected by Inversion : Behavioral and Electrophysiological Evidence. *Frontiers in Behavioral Neuroscience*, *11*, 21.

Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex. *Human brain mapping*, *27*(1), 77-89.

Srinivasan, N., & Hanif, A. (2010). Global-happy and local-sad : Perceptual processing affects emotion identification. *Cognition and Emotion*, *24*(6), 1062-1069.

Stuit, S. M., Kootstra, T. M., Terburg, D., van den Boomen, C., van der Smagt, M. J., Kenemans, J. L., & Van der Stigchel, S. (2021). The image features of emotional faces that predict the initial eye movement to a face. *Scientific Reports*, *11*(1), 8287.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature neuroscience*, *11*(7), 843-850.

Svard, J., Wiens, S., & Fischer, H. (2012). Superior recognition performance for happy masked and unmasked faces in both younger and older adults. *Frontiers in Psychology*, *3*, 520.

Sweeny, T. D., Suzuki, S., Grabowecky, M., & Paller, K. A. (2013). Detecting and categorizing fleeting emotions in faces. *Emotion*, *13*(1), 76-91.

Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., & de Gelder, B. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proceedings of the National Academy of Sciences*, *106*(42), 17661-17666.

Tamietto, M., Cauda, F., Corazzini, L. L., Savazzi, S., Marzi, C. A., Goebel, R., Weiskrantz, L., & de Gelder, B. (2010). Collicular vision guides nonconscious behavior. *Journal of cognitive neuroscience*, *22*(5), 888-902.

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, *11*(10), 697-709.

Tamietto, M., Pullens, P., de Gelder, B., Weiskrantz, L., & Goebel, R. (2012). Subcortical connections to human amygdala and changes following destruction of the visual cortex. *Current Biology*, *22*(15), 1449-1455.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology*, *46*(2), 225-245.

Tanaka, J. W., & Simonyi, D. (2016). The "parts and wholes" of face recognition : A review of the literature. *Quarterly Journal of Experimental Psychology*, *69*(10), 1876-1889.

Tannert, S., & Rothermund, K. (2020). Attending to emotional faces in the flanker task : Probably much less automatic than previously assumed. *Emotion*, *20*(2), 217.

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception : Evidence from the face inversion effect. *Vision research*, *51*(11), 1273-1278.

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology*, *29*, 97-101.

Tolhurst, D., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, *12*(2), 229-232.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242-249.

Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience*, *13*(2), 256-271.

Trautmann, S. A., Fehr, T., & Herrmann, M. (2009). Emotions in motion : dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain research*, *1284*, 100-115.

Treisman, A. (1985). Preattentive processing in vision. *Computer vision, graphics, and image processing*, *31*(2), 156-177.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annu. Rev. Neurosci.*, *31*, 411-437.

Tsuchiya, N., Moradi, F., Felsen, C., Yamazaki, M., & Adolphs, R. (2009). Intact rapid detection of fearful faces in the absence of the amygdala. *Nature neuroscience*, *12*(10), 1224-1225.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273-289.

Ungerleider, L. G., & Haxby, J. V. (1994). 'What'and 'where'in the human brain. *Current opinion in neurobiology*, *4*(2), 157-165.

Valentine, T. (1988). Upside-down faces : A review of the effect of inversion upon face recognition. *British journal of psychology*, *79*(4), 471-491.

Van der Schaaf, v. A., & van Hateren, J. v. (1996). Modelling the power spectra of natural images : statistics and information. *Vision research*, *36*(17), 2759-2770.

Van der Stigchel, S., & de Vries, J. P. (2015). There is no attentional global effect: attentional shifts are independent of the saccade endpoint. *Journal of Vision*, *15*(15), 17.

Van der Stigchel, S., & Nijboer, T. (2013). How global is the global effect? the spatial characteristics of saccade averaging. *Vision Research*, *84*, 6-15.

van der Gaag, C., Minderaa, R. B., & Keysers, C. (2007). The BOLD signal in the amygdala does not differentiate between dynamic facial expressions. *Social cognitive and affective neuroscience*, *2*(2), 93-103.

van der Schaaf, A., & van Hateren, J. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Research*, *36*(17), 2759-2770.

Van Le, Q., Isbell, L. A., Matsumoto, J., Nguyen, M., Hori, E., Maior, R. S., Tomaz, C., Tran, A. H., Ono, T., & Nishijo, H. (2013). Pulvinar neurons reveal neurobiological evidence of past selection for rapid detection of snakes. *Proceedings of the National Academy of Sciences*, *110*(47), 19000-19005.

VanRullen, R. (2006). On second glance : Still no high-level pop-out effect for faces. *Vision research*, *46*(18), 3017-3027.

Vassilev, A., & Mitov, D. (1976). Perception time and spatial frequency. *Vision research*, *16*(1), 89-92.

Veale, R., Hafed, Z. M., & Yoshida, M. (2017). How is visual salience computed in the brain? insights from behaviour, neurobiology and modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160113.

Victeur, Q., Huguet, P., & Silvert, L. (2019). Attentional allocation to task-irrelevant fearful faces is not automatic : Experimental evidence for the conditional hypothesis of emotional selection. *Cognition and Emotion*.

Vlamings, P. H. J. M., Goffaux, V., & Kemner, C. (2009). Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information? *Journal of Vision*, *9*(5), 12-12.

Vuilleumier, P. (2000). Faces call for attention : evidence from patients with visual extinction. *Neuropsychologia*, *38*(5), 693-700.

Vuilleumier, P. (2002). Facial expression and selective attention. *Current Opinion in Psychiatry*, *15*(3), 291-300.

Vuilleumier, P. (2015). Affective and motivational control of vision. *Current opinion in neurology*, *28*(1), 29-35.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, *6*(6), 624-631.

Vuilleumier, P., & Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception : evidence from functional neuroimaging. *Neuropsychologia*, *45*(1), 174-194.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature neuroscience*, *7*(11), 1271-1278.

Vuilleumier, P., & Sagiv, N. (2001). Two eyes make a pair : facial organization and perceptual learning reduce visual extinction. *Neuropsychologia*, *39*(11), 1144-1149.

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions : a voxel-based meta-analysis. *Journal of cognitive neuroscience*, *22*(12), 2864-2885.

Walker, R., & McSorley, E. (2006). The parallel programming of voluntary and reflexive saccades. *Vision Research*, *46*(13), 2082-2093.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366-383.

Wang, S., Xu, J., Jiang, M., Zhao, Q., Hurlemann, R., & Adolphs, R. (2014). Autism spectrum disorder, but not amygdala lesions, impairs social attention in visual search. *Neuropsychologia*, *63*, 259-274.

Ward, R., Calder, A. J., Parker, M., & Arend, I. (2007). Emotion recognition following human pulvinar damage. *Neuropsychologia*, *45*(8), 1973-1978.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. how individual face parts contribute to successful emotion recognition (M. A. Pavlova, Éd.). *PLOS ONE*, *12*(5), e0177239.

Weiskrantz, L., Warrington, E. K., Sanders, M., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain*, *97*(1), 709-728.

Weymar, M., & Schwabe, L. (2016). Amygdala and emotion : the bright side of it. *Frontiers in neuroscience*, *10*, 224.

Whalen, P. J., Kagan, J., Cook, R. G., Davis, F. C., Kim, H., Polis, S., McLaren, D. G., Somerville, L. H., McLean, A. A., Maxwell, J. S., et al. (2004). Human amygdala responsivity to masked fearful eye whites. *Science*, *306*(5704), 2061-2061.

Whalen, P. J., Raila, H., Bennett, R., Mattek, A., Brown, A., Taylor, J., van Tieghem, M., Tanner, A., Miner, M., & Palmer, A. (2013). Neuroscience and facial expressions of emotion : The role of amygdala–prefrontal interactions. *Emotion Review*, *5*(1), 78-83.

White, B. J., Berg, D. J., Kan, J. Y., Marino, R. A., Itti, L., & Munoz, D. P. (2017). Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nature communications*, *8*(1), 1-9.

White, B. J., Kan, J. Y., Levy, R., Itti, L., & Munoz, D. P. (2017). Superior colliculus encodes visual saliency before the primary visual cortex. *Proceedings of the National Academy of Sciences*, *114*(35), 9451-9456.

White, B. J., & Munoz, D. P. (2011). The superior colliculus.

Wiesmann, S. L., Caplette, L., Willenbockel, V., Gosselin, F., & Võ, M. L.-H. (2021). Flexible time course of spatial frequency use during scene categorization. *Scientific Reports*, *11*(1), 1-13.

Wilhelm, O., Hildebrandt, A., Manske, K., Schacht, A., & Sommer, W. (2014). Test battery for measuring the perception and recognition of facial expressions of emotion. *Frontiers in psychology*, *5*, 404.

Williams, M. A., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2005). Differential amygdala responses to happy and fearful facial expressions depend on selective attention. *Neuroimage*, *24*(2), 417-425.

Wollenberg, L., Deubel, H., & Szinte, M. (2018). Visual attention is not deployed at the endpoint of averaging saccades. *PLoS biology*, *16*(6), e2006548.

Wu, D. W.-L., Bischof, W. F., & Kingstone, A. (2014). Natural gaze signaling in a social context. *Evolution and Human Behavior*, *35*(3), 211-218.

Xu, P., Peng, S., Luo, Y.-j., & Gong, G. (2021). Facial expression recognition : A meta-analytic review of theoretical models and neuroimaging evidence. *Neuroscience & Biobehavioral Reviews*, *127*, 820-836.

Yarbus, A. L. (1967). *Eye movements and vision*. Springer.

Yiend, J. (2010). The effects of emotion on attention: a review of attentional processing of emotional information. *Cognition & Emotion*, *24*(1), 3-47.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, *81*(1), 141.

Yoonessi, A., & Yoonessi, A. (2011). Functional assessment of magno, parvo and koniocellular pathways ; current state and future clinical applications. *Journal of ophthalmic & vision research*, *6*(2), 119.

Young, A. W., Hellawell, D., & Hay, D. C. (2013). Configurational information in face perception. *Perception*, *42*(11), 1166-1178.

Yovel, G. (2016). Neural and cognitive face-selective markers : An integrative review. *Neuropsychologia*, *83*, 5-13.

Zelinsky, G., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the new York Academy of Sciences*, *1339*(1), 154.

Zelinsky, G., Zhang, W., Yu, B., Chen, X., & Samaras, D. (2006). The role of top-down and bottom-up processes in guiding eye movements during visual search. *Advances in neural information processing systems*, 1569-1576.

Zhang, X., Zhaoping, L., Zhou, T., & Fang, F. (2012). Neural activities in V1 create a bottom-up saliency map. *Neuron*, *73*(1), 183-192.

Zhaoping, L. (2002). A saliency map in primary visual cortex (TRENDS in Cognitive Sciences Vol.6 No.1.).

Zhou, N., Masterson, S. P., Damron, J. K., Guido, W., & Bickford, M. E. (2018). The mouse pulvinar nucleus links the lateral extrastriate cortex, striatum, and amygdala. *Journal of Neuroscience*, *38*(2), 347-362.