UNIVERSITE BRETAGNE LOIRE

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Romain COHENDET

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université de Nantes
sous le sceau de l'Université Bretagne Loire

École doctorale : Sciences et technologies de l'information, et mathématiques

Discipline : Informatique et applications, section CNU 27, 61, 16
Unité de recherche : Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN)

Soutenue le 12 décembre 2016

Prédiction computationnelle de la mémorabilité des images : vers une intégration des informations extrinsèques et émotionnelles

JURY

Présidente :	**M^me^ Alice CAPLIER**, Professeur des universités, Grenoble INP
Rapporteurs :	**M. Olivier LE MEUR**, Maître de conférences HDR, IRISA **M. Charles TIJUS**, Professeur des universités, Université Paris 8
Examinateurs :	**M. Matthieu PERREIRA DA SILVA**, Maître de conférences, Université de Nantes **M^me^ Anne-Laure GILET**, Maître de conférences, Université de Nantes
Invités :	**M. Vincent COURBOULAY**, Maître de conférences HDR, Université de La Rochelle **M. Mohammad SOLEYMANI**, NSF Ambizione fellow, Swiss Center for Affective Sciences
Directeur de thèse :	**M. Patrick LE CALLET**, Professeur des universités, Université de Nantes

Remerciements

Je tiens, tout d'abord, à remercier l'ensemble des membres du jury pour avoir laissé mon travail les concerner. Merci, en particulier, à Olivier Le Meur et Charles Tijus pour avoir accepté de relire cette thèse et d'en être les rapporteurs : vos remarques pertinentes n'ont pas laissé de jouer avec mon impatience de vous répondre. Merci également à Alice Caplier pour avoir présidé le jury avec finesse et détermination.

Il y a trois ans, j'ai quitté ma terre natale pour un ciel nébuleux. Merci à mes encadrants pour m'avoir montré l'esquisse d'un chemin – à Patrick, pour la lumière jetée en avant ; à Matthieu et Anne-Laure, pour avoir corrigé et affermi mon pas. L'horizon que nous avons levé ensemble aura, je l'espère, quelque attraction pour qui partage notre intérêt pour la mémorabilité et sa prédiction. Vous avez su tenir patience jusqu'à ce que l'urgence de ma besogne m'interdise de la délaisser ; à ma décharge, le mur du temps est bien relatif : il suffit d'une velléité pour le franchir. Succombant à vos conseils, j'ai dépixélisé mon discours jusqu'aux articulations, comme je sais combien le caractère d'un scientifique répugne aux genres de fantaisie dont un autre mêlerait ses phrases : je me suis mis à l'orthodoxie du style, qui prévaut dans cet ouvrage.

Merci également à toute l'équipe IVC pour m'avoir fait apprécier, lorsque le matin s'évanouissait dans la caféine, même ce plus fidèle ennemi ! Aux permanents : Nicolas, Vincent, JPEG, Benoit, Marcus, Harold, Christian et Toinon (je compte sur toi pour donner à Jean-Pierre des mots cools, mais ils rentrent pas dans la grille, alors il bondit de sa chaise :) Aux non permanents : Fillippo, Dimitri, Josselin, Alex, Lukas, Yoann, Karam, Ahmed, Yashas, Ting, Suiyi, Geoffrizzz, Julius, Ervan, Chen...

Merci, bien sûr, aux amis de chez moi – en particulier, pour les théories composées à nos heures dérobées et les souvenirs de montagne – les montagnes sont souvent venues me chercher à Nantes, et mes obsessions de la verticalité et des grands espaces sont intactes. Fabrice, Jérèm, Tanoh, Zuska, Sylvain et Matthieu : on se croise bientôt au château ; Nabil hante déjà le mobilier – ou pas... il en sera sorti.

Merci, enfin, à ma famille : j'aurais pu vendre des sûretés, mettre les chiffres des autres dans des lignes ou laïusser pour trouver grâce aux yeux de tous, les vôtres y auraient trouvé de quoi étinceler.

En somme, merci à tous, pour avoir fait, lorsque la thèse a resserré mon monde, qu'il se soit réduit à votre société.

Table des matières

Liste des tableaux

Table des figures

Introduction générale

Contexte et motivation

La révolution numérique a eu un impact considérable sur la quantité d'images numériques partagées quotidiennement. Accompagnant la diffusion massive des écrans et des appareils photographiques, les images numériques se multiplient dans la plupart des domaines de la vie quotidienne. Selon l'entreprise Facebook, en 2013, 350 millions de photos inédites ont été téléversées chaque jour sur le site éponyme, pour ne parler que de celui-ci (Ericsson and Qualcomm, 2013). Google assurait, en 2008, avoir indexé plus de 1000 milliards d'images (Google, 2008). Il n'est pas étonnant que le monde numérique s'augmente d'une telle quantité d'images, étant donnée l'importance de la vision chez l'être humain. On estime ainsi que près d'un tiers de notre cerveau participerait au traitement de l'information visuelle (Chalupa and Werner, 2004). Cependant, se pose la question de l'organisation de cette multitude d'images, avec pour finalité de proposer une expérience de qualité à un utilisateur donné, dans un contexte donné.

L'évolution nous a doués d'une mémoire remarquable pour les images, qui nous permet de reconnaître des milliers d'images que nous n'avons vues qu'une seule fois (Standing, 1973). Cependant, si certaines images nous marquent presque indélébilement, nous sommes incapables d'en reconnaître d'autres que nous avons vues, pourtant, quelques minutes auparavant seulement (Isola et al., 2011b). On peut penser que notre cerveau, si l'évolution l'a fait efficace, tend à mémoriser les images qui nous sont utiles. Dans ce sens, la mémorabilité d'une image[1] nous dit quelque chose de son importance dans la vie quotidienne, comme cela a été proposé par (Isola et al., 2014). Elle pourrait ainsi nous permettre de mieux organiser le monde des images. Imaginons qu'on veuille choisir dans une base de données une image pour un utilisateur particulier, ou créer un matériel d'éducation, un site internet, une interface utilisateur, etc. ; ou encore résumer une vidéo par une seule image, choisir pour un réseau social un de nos selfies... Il est de nombreuses situations où la mémorabilité d'une image est une propriété qu'il est hautement désirable de maximiser.

[1] Dans son acception la plus générique, telle qu'elle subsume l'ensemble des définitions auxquelles nous nous référons dans cette thèse, la mémorabilité d'une image renvoie à la probabilité qu'elle soit mémorisée alors qu'elle est vue une certaine quantité de temps par un individu.

La prédiction automatique de la mémorabilité des images est un sujet de recherche récent (Isola et al., 2011b), qui a rapidement suscité l'intérêt des chercheurs en vision par ordinateur (p. ex.(Khosla et al., 2012b,Mancas and Le Meur, 2013,Oliva et al., 2013,Bylinskii et al., 2015b,Celikkale et al., 2015,Wang et al., 2015,Lahrache et al., 2016)). Les première tentatives ont reposé sur l'utilisation d'algorithmes d'apprentissage pour inférer de caractéristiques de bas niveau des images leur degré de mémorabilité. Elles ont montré qu'il est, dans une certaine mesure, possible de prédire computationnellement le degré de mémorabilité d'une image. Cependant, ce champ de recherche en est à ses débuts, et de nombreuses possibilités pour améliorer les prédictions n'ont pas encore été exploitées.

Approche et contributions

Le premier article traitant, à notre connaissance, de la mémorabilité des images en informatique date de 2011 (Isola et al., 2011b). Dans cet article fondateur, Isola *et al.* définissent la mémorabilité d'une image comme la probabilité qu'elle soit reconnue après un délai de rétention mnésique de quelques minutes. Ils proposent également une base de 2222 images associées à des scores de mémorabilité. Cette définition de la mémorabilité sera reprise par les études subséquentes qui s'inscrivent dans la lignée des travaux de ces auteurs (Isola et al., 2011a,Khosla et al., 2012a,Khosla et al., 2012b,Khosla et al., 2013, Mancas and Le Meur, 2013, Kim et al., 2013, Celikkale et al., 2013, Oliva et al., 2013, Isola et al., 2014, Bylinskii et al., 2015b, Bylinskii et al., 2015a, Celikkale et al., 2015, Wang et al., 2015, Lahrache et al., 2016), sans que le concept de mémorabilité ne soit fondamentalement mis en lien avec la littérature sur la mémoire, abondante en psychologie, où elle est étudiée depuis plus d'un siècle. De plus, la base de données proposée par Isola *et al.* — la seule disponible à ce jour —, servira de matière pour entraîner les modèles dont nous avons connaissance, ce qui pose problème pour l'évaluation de la capacité de ces modèles à généraliser.

Que prédisent, en fait, les modèles existants ? Des questions à nos yeux essentielles n'ont pas été posées ; par exemple, concernant la durée de la rétention mnésique où l'émotion véhiculée par les images. Quels sont les biais potentiels susceptibles d'avoir influencé la mesure de mémorabilité ? La qualité de la *vérité terrain* utilisée pour l'apprentissage et l'évaluation des différents modèles est un point essentiel dans la création de métriques objectives (Krig, 2014). Pour bien répondre à ces différentes questions, il est nécessaire d'intégrer l'étude de mémorabilité des images dans une approche théorique clairement définie.

A ce jour, l'étude de la mémorabilité des images en vision par ordinateur aura principalement consisté à découvrir des caractéristiques spécifiques aux images mémorables, dans l'objectif de développer un modèle représentatif de ce type d'image. Cette approche a permis d'obtenir des résultats, cependant modestes. A titre indicatif, la per-

formance du modèle de (Isola et al., 2014), déterminée par le coefficient de corrélation des rangs de Spearman entre les scores de mémorabilité *vérité terrain* et les scores de mémorabilité prédits, est de .462, avec une erreur quadratique moyenne de .017. Le comblement de certaines lacunes dans les travaux existants pourrait améliorer substantiellement la performance des modèles de prédiction. En particulier, les informations extrinsèques de l'image, relatives au contexte de présentation de l'image et aux individus qui les regardent, ont été à peine exploitées ; et l'apprentissage profond (en anglais, *Deep learning*), dont les preuves de l'efficacité pour traiter des problèmes similaires sont nombreuses, n'a pas été essayé pour la prédiction de la mémorabilité[2].

D'autre part, il a été montré que les modèles de saillance, qui permettent de révéler dans l'image les zones les plus susceptibles d'attirer l'attention humaine, et de calculer à partir des cartes de saillance qu'ils génèrent des caractéristiques de l'image liées à l'attention visuelle, sont des outils intéressants pour la prédiction computationnelle de la mémorabilité (Mancas and Le Meur, 2013). Selon Mancas et Le Meur, l'attention visuelle devrait être davantage considérée, en lien avec les caractéristiques de bas niveau des images, dans le cadre de la prédiction de la mémorabilité d'images.

L'approche adoptée dans cette thèse est transversale : elle fait intervenir des concepts et des outils de psychologie aussi bien que d'informatique. Nous revenons sur les fondements théoriques de la mémorabilité des images et insistons particulièrement sur les émotions véhiculées par les images, qui sont étroitement liées à leur mémorabilité. En considération de cet éclairage théorique, nous proposons d'inscrire la prédiction de la mémorabilité des images dans un cadre de travail plus large, qui embrasse les informations intrinsèques et extrinsèques (contextuelles et individuelles).

En conséquence, nous construisons une nouvelle base de données pour l'étude de la mémorabilité d'images. Elle sera utile pour éprouver les modèles existants, entraînés sur l'unique base de données existante, proposée par (Isola et al., 2011b). Notre base de données comprend 150 images associées à des scores de mémorabilité et d'émotion, pour lesquelles nous avons également collecté les données oculométriques des observateurs lors de leur visionnage. A partir de cette matière, nous étudions les liens entre l'émotion véhiculée par les images et leur mémorabilité, et nous intéressons à la question de la diminution de la mémorabilité des images en mémoire à long terme.

Nous introduisons également l'apprentissage profond pour la prédiction de la mémorabilité d'image : notre modèle obtient les meilleurs résultats à ce jour. Pour améliorer cette prédiction, nous investiguons plusieurs pistes pour prendre en compte les effets du contexte (notamment, émotionnel) de présentation des images dans la prédiction de la mémorabilité. Nous cherchons également une voie qui permette d'intégrer

[2]Ou plus exactement, n'avait pas encore été essayé : au moment où nos travaux sur l'apprentissage profond ont été réalisés, Khosla *et al.* ont développé une approche similaire. L'article présentant ces travaux (Khosla et al., 2015) n'était cependant pas encore disponible.

l'idiosyncrasie[3] dans l'équation. Conséquemment, nous proposons un modèle des liens entre facteurs individuels et probabilité de reconnaître une image répétée.

En outre, nous étudions le lien entre attention visuelle, émotion et mémorabilité : nos résultats confirment l'intérêt d'utiliser des caractéristiques de l'image liées à l'attention visuelle pour la prédiction de la mémorabilité. Nous éprouvons également plusieurs modèles de saillance sur notre base de données, et testons si leur performance varie en fonction du degré de mémorabilité des images et de l'émotion qu'elles véhiculent. Nos résultats montrent que certains modèles présentent effectivement un bais.

Finalement, nous proposons un outil, le « film interactif émotionnel », dont le déroulement dépend des réactions émotionnelles du spectateur. L'émotion est mesurée en temps réel, à l'aide d'un dispositif EEG qui fournit directement des données émotionnelles, en même temps qu'un oculomètre enregistre les mouvements oculaires du spectateur. Cet outil possède un potentiel intéressant pour élargir nos travaux sur la mémorabilité aux vidéos, question qui ne saurait tarder de susciter l'intérêt de notre communauté.

Organisation de la thèse

La suite de cette thèse est organisée comme suit (voir la figure 1).

Dans la **PARTIE I**, nous rapprochons l'étude théorique de la mémoire en psychologie de celle de la mémorabilité des images en informatique. Le **chapitre 1** porte sur la définition, les catégorisations et les mesures de la mémoire humaine. Le **chapitre 2** traite de la définition et de la mesure et des émotions, qui jouent un rôle essentiel dans les processus mnésiques. Le **chapitre 3** porte sur les liens entre émotion et mémoire, et ce qu'ils impliquent pour l'étude de la mémorabilité des images. Le **chapitre 4** consiste en une synthèse des travaux portant sur la prédiction de la mémorabilité des images en informatique. Dans le **bilan** de la partie I, nous proposons un résumé des lacunes dans notre domaine d'étude révélées par notre étude de la littérature, et présentons les moyens employés dans la suite de la thèse pour y remédier.

La **PARTIE II** porte sur la création d'une base de données pour l'étude de la mémorabilité des images, et sur l'analyse des scores de mémorabilité et d'émotion des images de cette base. Dans le **chapitre 5**, nous détaillons l'expérience mise en place pour obtenir des scores d'émotion et de mémorabilité, ainsi que les données oculométriques des participants, pour 150 images. Le **chapitre 6** traite de l'analyse des scores d'émotion obtenus dans le chapitre précédent, et de leur mise en relation avec les scores obtenus dans les études précédentes. Le **chapitre 7** porte sur l'analyse conjointe des scores d'émotion et de mémorabilité. Nous nous y intéressons aux effets de l'émotion

[3] Selon le dictionnaire Larousse, l'idiosyncrasie correspond à « la manière d'être particulière à chaque individu qui l'amène à avoir tel type de réaction, de comportement qui lui est propre. »

sur la mémorabilité des images, et investiguons la question du rôle de l'émotion dans l'évolution en mémoire à long terme de la mémorabilité des images.

La **PARTIE III** est consacrée à l'approche que nous avons mise en place pour répondre au défi de la prédiction de la mémorabilité d'images. Dans le **chapitre 8**, nous introduisons l'apprentissage profond pour la prédiction de la mémorabilité d'images. Dans le **chapitre 9**, nous explorons différentes voies pour prendre en compte le contexte de présentation d'une image (en particulier, le contexte émotionnel) dans la prédiction computationnelle de la mémorabilité d'images. Dans le **chapitre 10**, nous proposons un modèle de l'influence de plusieurs facteurs individuels sur la probabilité de reconnaître une image. L'objectif est de mettre en lumière les informations individuelles qui pourraient être prise en compte pour personnaliser les scores de mémorabilité prédits par les modèles.

La **PARTIE IV** porte sur la modélisation de l'attention visuelle pour l'étude de la mémorabilité des images et de l'émotion qu'elles véhiculent. Nous y proposons également un outil que nous avons développé, qui présente un potentiel intéressant pour élargir nos travaux aux vidéos. Dans le **chapitre 11**, nous confirmons le lien entre, d'un côté, l'attention visuelle, et de l'autre la mémorabilité des images et l'émotion qu'elles véhiculent, à partir de l'analyse des données oculométriques enregistrées durant l'expérience présentée dans le chapitre 5. Nous comparons également la performance de plusieurs modèles computationnels de saillance, et testons si cette performance varie selon le degré de mémorabilité des images et l'émotion qu'elles véhiculent. Dans le **chapitre 12**, nous présentons le « film interactif émotionnel », dont le fonctionnement repose sur l'enregistrement en temps réel des émotions du spectateur à l'aide d'un dispositif EEG qui fournit des données émotionnelles.

Nous terminons cette thèse par une **conclusion générale**. Nous y fournissons une synthèse du travail réalisé, et proposons des perspectives pour nos travaux futurs.

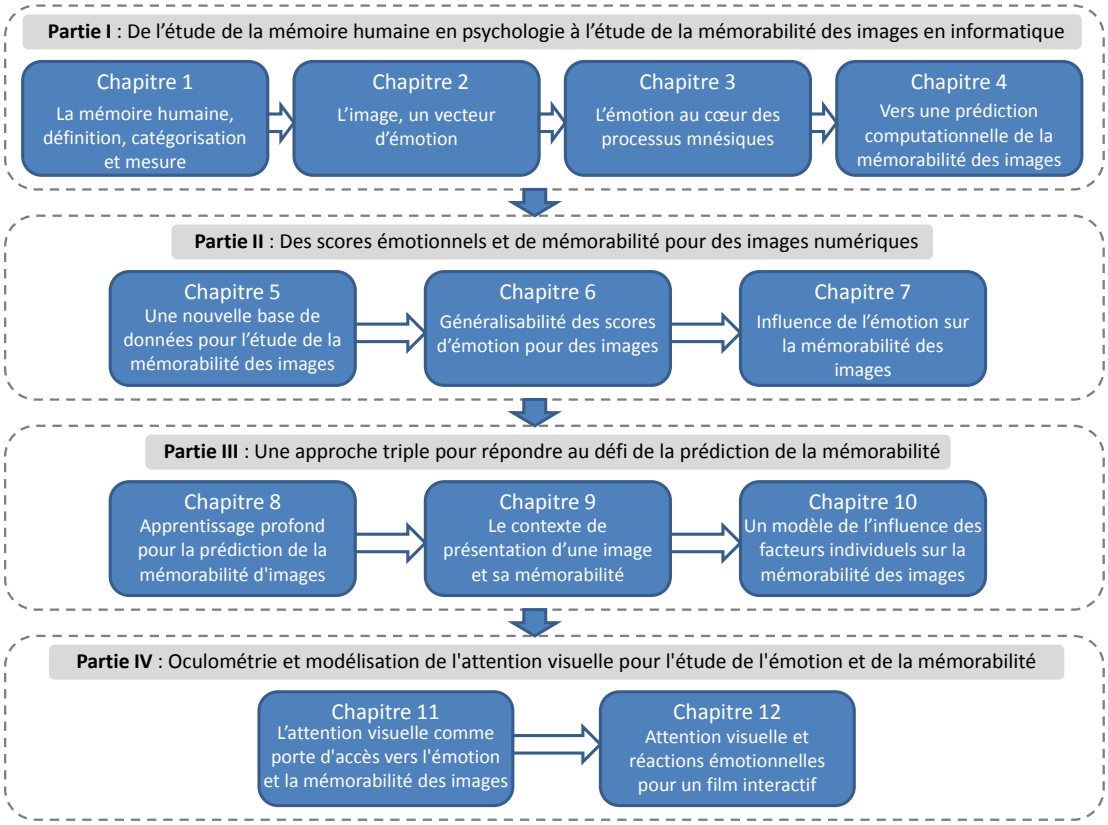


FIGURE 1 – Schéma du plan de thèse.

De l’étude de la mémoire humaine en psychologie à l’étude de la mémorabilité des images en informatique

Introduction

La mémorabilité des images est un sujet d'étude récent en informatique. La mémoire est cependant étudiée depuis plus d'un siècle en psychologie, et l'essentialité de sa relation avec l'émotion n'est plus à démontrer. Dans cette première partie, nous faisons le lien entre l'étude de la mémorabilité des images en informatique et celle de la mémoire en psychologie. L'objectif n'est pas de dresser un état de l'art exhaustif de la mémoire humaine et de l'émotion. Notre volonté est de produire une vue d'ensemble, synthétique, qui réponde aux objectifs suivants : renforcer l'assise théorique du champ de recherche dans lequel nous inscrivons nos travaux, dont l'objet est la mémorabilité des images ; mettre en lumière certaines lacunes dans les travaux existants, qui décèlent des leviers actionnables pour augmenter la performance des modèles de prédiction ; expliciter les concepts auxquels il sera fait référence dans cette thèse ; présenter les outils dont nous nous servirons.

Le chapitre 1 porte sur la définition, les catégorisations et les mesures de la mémoire humaine, objet d'étude qui suscite un intérêt important de la communauté scientifique depuis la révolution cognitive au milieu du siècle précédent (qui donnera naissance aux sciences cognitives), et qui, par conséquent, a conduit à la production d'une littérature considérable.

Le chapitre 2 porte spécifiquement sur les émotions humaines. Ce chapitre nous est apparu nécessaire étant donnée l'importance du rôle joué par l'émotion dans les processus mnésiques. De surcroît, la recherche sur l'émotion a récemment connu un élan considérable en informatique (imprimé notamment par la montée en puissance de l'informatique affective (Picard and Picard, 1997, Picard, 2010)), et un champ de recherche porte spécifiquement sur la prédiction computationnelle de l'émotion véhiculée par les images (p. ex. (Liu et al., 2010b, Wei et al., 2008, Gbèhounou et al., 2012, Machajdik and Hanbury, 2010)), ce qui pourrait faciliter le rapprochement de l'étude de l'émotion et de la mémorabilité des images en informatique.

Dans le chapitre 3, nous étudions les liens entre les processus émotionnels et mnésiques. En particulier, nous documentons l'amélioration de la mémoire des images par l'émotion qu'elles véhiculent. Puis nous expliquons que la mémorabilité des images est susceptible de varier dans le temps, en raison du processus de consolidation des souvenirs en mémoire à long terme, et que l'émotion joue un rôle essentiel dans ce

phénomène.

Le chapitre 4 traite de l'étude de la mémorabilité des images en informatique. Les études existantes ont principalement porté sur la recherche de caractéristiques spécifiques aux images mémorables, dans l'objectif de développer un modèle représentatif de ce type d'images. Les informations liées au contexte de présentation de l'image et aux observateurs n'ont pratiquement pas été considérées. Nous reviendrons sur les différentes méthodes employées, et soulignerons celles qui auraient pu l'être.

Nous terminons cette première partie par un bilan, qui éclaire certaines lacunes de la recherche sur la mémorabilité des images en informatique à la lumière des théories sur la mémoire et l'émotion développées en psychologie.

1

La mémoire humaine : définition, catégorisation et mesure

La prédiction de la mémorabilité d'une image est un problème d'informatique : il s'agit de lier des informations extraites de l'image de manière computationnelle à des scores de mémorabilité associés à ces images. Cependant, pour comprendre à quoi correspond la mémorabilité que l'on prédit, il est nécessaire de comprendre les phénomènes de mémoire qui ont été à l'œuvre lors de la mesure de cette mémorabilité et les facteurs qui ont pu l'influencer.

1.1 Définition de la mémoire

L'utilisation d'un terme unique — la mémoire — suggère l'existence d'un système unitaire. Pourtant, après plus de 2000 ans de spéculations philosophiques et plus d'un siècle d'étude scientifique (Baddeley, 1997), la conception en systèmes multiples de la mémoire prime désormais sur la conception unitaire. Il aura fallu attendre les années 1950 et la fin du behaviorisme dominant, portant la conception d'une mémoire constituée d'un seul réseau associatif, pour que les chercheurs, influencés par le traitement de l'information, commencent à distinguer des systèmes de mémoire différents (Lieury, 2005). Ces systèmes varient notamment, comme nous allons le voir, selon la durée de stockage de l'information (de quelques millisecondes pour le registre sensoriel (Sperling, 1960) à la vie entière pour la mémoire à long terme), ou encore selon la nature de la mémoire ou de sa manifestation (p. ex. (Tulving, 1972, Schacter, 1985)). Au-delà

d'une réflexion par systèmes, lorsque son étude s'inscrit dans une perspective de traitement de l'information, essentielle dans le courant cognitiviste classique, la mémoire se rapporte aux processus d'encodage, de stockage et de récupération des représentations mentales. L'encodage renvoie au processus de traitement de l'information qui vise à transformer les informations pour les rendre compatibles avec le système mnésique. Il correspond à la phase d'acquisition des informations. Le stockage renvoie au processus de conservation des informations. Il correspond à la phase de rétention des informations. La récupération correspond à la phase de restitution des informations.

1.1.1 Catégorisation fondée sur la durée de la mémoire

Le modèle modal

Parmi les modèles structuraux de la mémoire, le modèle modal est certainement le plus influent. Sa formulation classique a été proposée en 1968 par Atkinson et Shiffrin (Atkinson and Shiffrin, 1968). Suivant celle-ci, la mémoire correspond à trois sous-systèmes, indépendants dans leur fonctionnement mais reliés : une mémoire sensorielle à très court terme (le registre sensoriel), une mémoire à court terme (MCT) et une mémoire à long terme (MLT). C'est un modèle de type sériel : les informations transitent d'abord par la mémoire sensorielle, puis par la MCT, puis parviennent à la MLT. Lors du passage dans la MCT, l'information peut faire l'objet d'une répétition articulatoire ou d'une élaboration (i.e. la création de liens faisant sens) permettant son passage en MLT, c'est-à-dire sa pérennisation.

Pour prendre l'exemple d'une image, si celle-ci est perçue, admettons, 200 ms, la mémoire iconique (i.e. la mémoire sensorielle visuelle) permet de retenir les informations visuelles pour une durée maximum de 500 ms (après la disparition de l'image). Puis l'information passe en MCT, ce qui s'accompagne d'une première perte de détails visuels ; elle y demeure jusqu'à quelques dizaines de secondes (cette durée dépend de l'auto-répétition de l'information en mémoire). Enfin, une partie de l'information passe en MLT, où n'est généralement conservée que l'essentiel de l'image (Brady et al., 2008). Le passage de l'information en MLT, où elle sera durablement stockée, dépend de plusieurs facteurs, au premier rang desquels l'émotion véhiculée par l'image et l'attention portée aux différentes informations — nous y reviendrons dans le chapitre 3 (section 3.1).

La mémoire sensorielle (ou registre sensoriel)

La mémoire sensorielle se situe à l'interface entre les sens et les systèmes mnésiques supérieurs. Elle dure quelques millisecondes, et conserve — le temps nécessaire à leur fixation en MCT — les stimuli sensoriels (sons, images, odeurs, etc.).

Sperling a fourni une preuve comportementale considérable de l'existence de la mémoire sensorielle (Sperling, 1960). Dans un premier temps, il a présenté à des participants des images comprenant 3 lignes de 4 lettres, pendant une période très brève (50 ms). Lorsqu'il a demandé aux participants de rappeler les items, ils ont rapporté en moyenne 4 à 5 lettres sur 12. Deux interprétations s'offraient alors à lui : soit les participants n'avaient pas eu le temps de voir toutes les lettres, et donc n'en rappelaient qu'un petit nombre, soit ils avaient vu toutes les lettres mais en oubliaient rapidement certaines. Si ce dernier cas était avéré, cela aurait supposé l'existence d'une mémoire à très court terme, dite sensorielle. Afin de trancher entre ces deux hypothèses, Sperling a réitéré son expérience, avec le même type de matériel, mais en précisant cette fois-ci aux participants qu'ils auraient à se souvenir seulement des lettres d'une ligne en particulier. Au moment du rappel, le participant était prévenu de la ligne à rappeler par un son : un son aigu pour la ligne supérieure, un son moyen pour la ligne médiane et un son grave pour la ligne inférieure. Dans cette condition, les participants étaient capables de rappeler 3 lettres par lignes. Ainsi, dans le premier cas les participants se rappelaient de 4 à 5 lettres en moyenne sur les 12 présentées, tandis que dans le deuxième cas ils se rappelaient de 3 lettres sur 4, soit — théoriquement — 9 sur 12. Sperling en a tiré la conclusion que, puisque les participants n'avaient aucun moyen de connaître à l'avance quelle ligne serait à rappeler, il était probable qu'au moment du rappel ils possédaient au moins 9 lettres dans leur mémoire sensorielle. Cependant, cette mémoire sensorielle ne durant qu'un laps de temps très court, les éléments ayant le temps d'être récupérés par cette méthode de rappel[1] étaient moins nombreux que la capacité réelle de stockage du système de mémoire.

Une dichotomie MCT/MLT

L'existence d'une MCT distincte de la MLT repose sur un certain nombre d'arguments, parmi lesquels ceux qui suivent sont les plus souvent invoqués.

Premièrement, cette distinction est étayée par le cas, célèbre, du patient H.M., décrit notamment par Milner en 1966 (Milner, 1966), qui a contribué significativement à la connaissance sur les mécanismes de la mémoire. H.M. souffrait de crises d'épilepsie pharmaco-résistantes : il a subi une opération chirurgicale visant à l'ablation bilatérale d'une partie importante des lobes temporaux et de l'hippocampe (structures neurales au cœur de la fonction de mémoire). A la suite de cette opération, il commença à souffrir d'amnésie antérograde, c'est-à-dire qu'il se trouva dans l'incapacité de fixer durablement de nouveaux souvenirs bien qu'il possédât encore toutes ses capacité de raisonnement et de MCT (ce qui a été vérifié en utilisant une tâche d'empan mnésique[2]). La

[1] Les différentes méthodes de mesure de la mémoire sont présentées dans la section 1.2.

[2] L'empan mnésique est une notion introduite par (Miller, 1956), qui désigne le nombre d'éléments qu'un individu peut restituer immédiatement après leur présentation, généralement en respectant l'ordre dans lequel ils ont été présentés.

dichotomie MCT/MLT rend compte de ce résultat : la MCT peut être préservée alors que la MLT est altérée.

A la suite du cas H.M., les chercheurs ont observé des patients cérébrolésés présentant un important déficit de MCT sans déficit associé de MLT (Shallice and Warrington, 1970). Ces observations remettent en cause la sérialité présente dans le modèle d'Atkinson et Shiffrin, selon laquelle la MLT se construit à partir de la MCT, qui elle-même se construit à partir des registres sensoriels. Cependant, elles renforcent en même temps l'idée qu'il existerait deux registres mnésiques différents — MCT et MLT —, en ceci qu'elles permettent de faire une double dissociation (un patient ayant la MCT fonctionnelle et la MLT lésée, et un autre patient ayant la MCT lésée et la MLT fonctionnelle).

L'effet de position sérielle, qui advient lors de la récupération de l'information stockée en mémoire, est également un argument avancé en faveur de l'existence d'une dichotomie MCT/MLT. Cette notion renvoie au fait que les taux de rappel des différents éléments d'une liste mémorisée varient en fonction de leur position dans la liste lorsque ce taux de rappel est mesuré par un test de mémoire immédiat (i.e. passé immédiatement après que la liste d'éléments a été présentée) (Murdock Jr, 1962). Ainsi, les premiers et les derniers éléments de la liste sont généralement les mieux rappelés ; on parle d'effets de primauté, et de récence. L'expérience classique pour montrer ces effets consiste à présenter à des participants un à un les éléments d'une liste, puis de leur faire passer un test de rappel libre immédiat. A partir des résultats, on peut alors tracer une courbe de position sérielle en faisant correspondre à chaque position d'un élément dans la liste un pourcentage moyen de rappel. La courbe obtenue montre généralement que les premiers et les derniers éléments sont mieux rappelés que les éléments situés au milieu de la liste, ce qui permet de constater les effets de primauté et de récence. La figure 1.1 présente une courbe de position sérielle hypothétique. On explique généralement l'effet de position sérielle par l'existence d'une dichotomie MCT/MLT. Au début de la liste, le nombre d'éléments à mémoriser étant faible, le passage des informations de la MCT à la MLT se fait facilement. A mesure que la quantité d'information à mémoriser augmente, la MCT est surchargée, et une partie de l'information maintenue en MCT est remplacée sans être passée en MLT. Quant aux derniers éléments de la liste, ils sont encore en MCT au moment du rappel immédiat, ce qui explique l'effet de récence.

Mémoires à court et long terme et mémorabilité des images

La durée de stockage des informations en MCT est de quelques dizaines de secondes au maximum (sans répétition de l'information) (Revlin, 2012). La mémorabilité d'une image telle qu'elle est étudiée en informatique, où elle correspond, d'après la définition de (Isola et al., 2011b), à « la probabilité que l'image soit reconnue après un délai de rétention mnésique de quelques minutes lorsqu'elle est présentée dans une suite images », renvoie donc à une performance de MLT. Cependant, en accord avec le modèle modal, l'information passe par la MCT avant de passer en MLT : une surcharge de la MCT

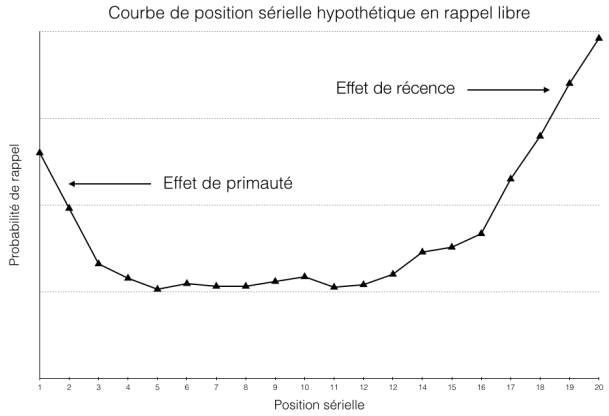


FIGURE 1.1 – Courbe de position sérielle hypothétique en rappel libre (adaptée de (Deese and Kaufman, 1957)).

est donc susceptible d'influencer négativement la mémorisation en MLT. Pour se rapprocher des conditions de mémorisations dans la vie quotidienne, où une personne peut généralement visionner autant de temps qu'elle le souhaite les images qui l'intéressent, il sera donc pertinent de ne pas surcharger la MCT des participants dans une tâche de mémoire visant à obtenir des scores de mémorabilité. En particulier, on veillera à fixer un SOA[3] suffisamment long pour ne pas saturer la MCT. La capacité de la MCT, de 7+-2 éléments chez l'adulte (ce chiffre a été obtenu en utilisant des tâches d'empan mnésique) (Miller, 1956), et sa durée d'environ 18 secondes (chez l'adulte, sans répétition de l'information) selon (Revlin, 2012), nous fournissent des indications pour fixer ce SOA : on choisira un SOA au minimum supérieur à $18/7$, soit 2.6 secondes entre deux images, pour éviter une surcharge de la MCT. Dans (Isola et al., 2011b), ce SOA est de 2,4 secondes. En considération de l'effet de primauté, on veillera également à intervertir l'ordre des images présentées aux différents participants (comme cela a été fait dans (Isola et al., 2011b)) pour éviter que les scores de mémorabilité des images ne dépendent de leurs positions dans la série d'images présentées.

Si la capacité de la MCT est relativement limitée, la capacité de la MLT est extrêmement vaste. C'est en particulier vrai pour la mémoire d'images, lorsqu'elles est mesurée par un test de reconnaissance (voir 1.2) (Standing, 1973). La préoccupation principale concernant la quantité d'items évalués par une tâche de mémoire destinée à mesurer le degré de mémorabilité d'images portera sur la fatigue des participants plutôt que sur leur capacité à mémoriser un grand nombre d'images.

[3]Le SOA, ou *Stimulus Onset Asynchronie*, correspond au temps qui sépare le début de la présentation d'un item du début de la présentation de l'item qui le suit.

1.1.2 Catégorisation fondée sur la nature de la mémoire

Outre la conception multi-systèmes de la mémoire fondée sur la durée, d'autres divisions ont été proposées pour la MCT (modèle de la mémoire de travail (Baddeley, 1986)) et la MLT (dichotomies mémoire épisodique/sémantique (Tulving, 1972), déclarative/procédurale (Cohen and Squire, 1980), explicite/implicite (Graf and Schacter, 1985)). La figure 1.2 présente une taxonomie des systèmes de mémoire communément admis.

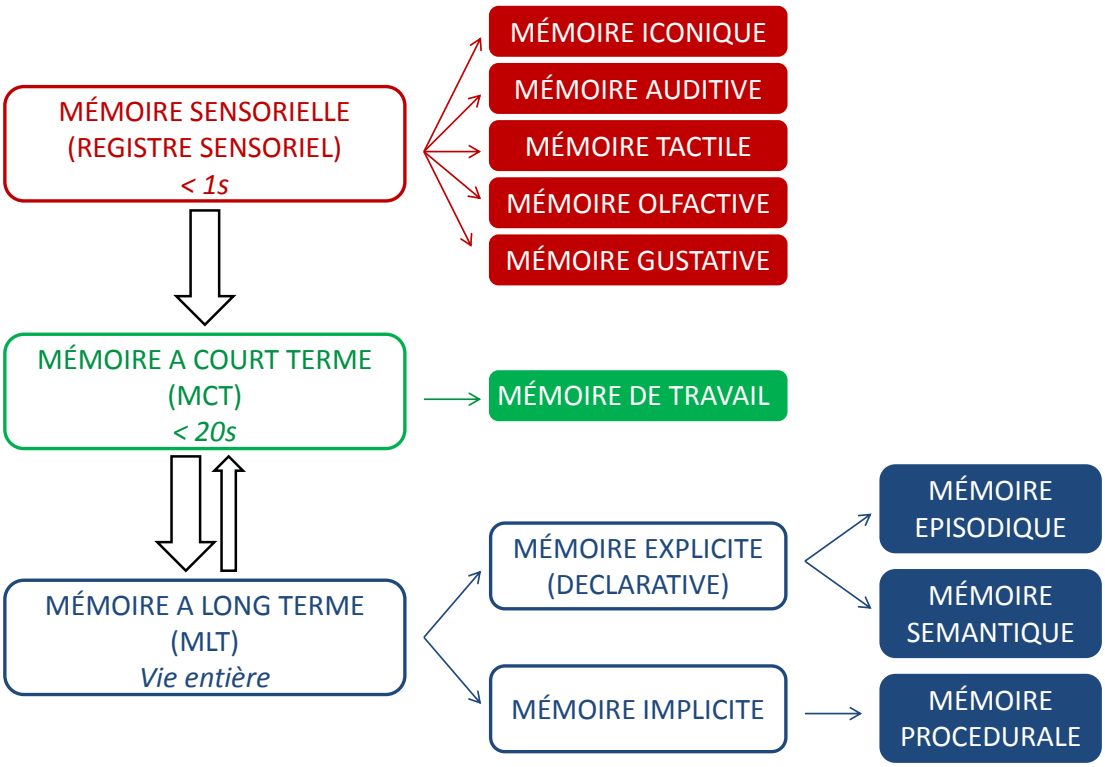


FIGURE 1.2 – Taxonomie des systèmes de mémoire communément admis.

Mémoire de travail

Après que la mémoire a été dissociée en court et long terme, l'hypothèse que la MCT était elle-même multiple a été soulevée, pour la première fois par Baddeley en 1986 (Baddeley, 1986). Il propose que la mémoire de travail serait composée d'un processeur central contrôlant deux systèmes esclaves : la « boucle phonologique » et le « calepin visuo-spatial ». Le premier système — la boucle phonologique — s'appuie sur les travaux antérieurs sur la mémoire lexicale, et intègre les stratégies de rétention de l'information lexicale (l'auto-répétition, qui peut significativement augmenter la durée de

rétention des informations en MCT, et la subvocalisation, qui a également un effet facilitateur sur la mémorisation (Slowiaczek and Clifton, 1980)). Le calepin visuo-spatial maintient l'information visuo-spatiale à court terme. Il permet la génération et la manipulation des images mentales. Baddeley et Hitch (1974) supposent qu'il est constitué de deux composantes : un registre de stockage, passif — c'est-à-dire un stock visuel à court terme, dans lequel l'information visuelle décline rapidement et est sensible à l'interférence — et un processus de rafraîchissement des informations. Le processeur central est un mécanisme attentionnel qui intègre les informations issues de la boucle phonologique et du calepin visuo-spatial et les mets en relation avec les connaissances stockées en MLT.

La mémoire de travail étant sévèrement limitée (Cowan, 2012), il est important, lorsqu'une tâche est proposée dans le cadre d'une étude scientifique, que les informations utiles pour accomplir cette tâche puissent entrer dans les limites de la mémoire de travail. La quantité totale d'effort mental utilisée par la mémoire de travail, ou charge cognitive (Sweller, 1988), est étroitement liée à la capacité de stockage d'informations en mémoire de travail. Si un trop grand nombre d'informations doit être traité simultanément pour réussir une tâche, la charge cognitive est alors trop élevée : la mémoire de travail est surchargée, ce qui entraine l'échec de la tâche ou, plus important en ce qui nous concerne, une mauvaise mémorisation en MLT. D'autant que, dans une tâche visant à mesurer une performance de MLT, une mauvaise performance due à une surcharge cognitive pourrait être prise à tort pour une mauvaise performance de MLT. Imaginons, par exemple, une tâche de mémoire ou des images seraient présentées très rapidement lors de la phase d'encodage, préalable à la phase de reconnaissance qui interroge la MLT. La performance de mémoire étant mesurée en MLT grâce aux résultats de reconnaissance, il sera difficile de dire si le nombre limité d'images reconnues est dû aux limitations de la MLT ou à celles du passage de l'information de la mémoire de travail à la MLT. Ce point doit attirer notre attention sur le fait que l'attribution d'une performance de mémoire à un système de mémoire particulier (si systèmes il existe) est chose difficile, comme de nombreux mécanismes sont souvent impliqués (par ex., l'attention visuelle, la préférence, la familiarité, etc.). Il s'agira donc de privilégier un raisonnement considérant des performances relatives plutôt qu'absolues. En particulier, la mémorabilité d'une image n'a de sens que comparativement à celle des autres images, en lien avec un contexte de mémorisation spécifique.

La dichotomie Mémoire épisodique/Mémoire sémantique

Cette division structuro-fonctionnelle primordiale de la MLT en mémoire épisodique et mémoire sémantique, a été initialement proposée en 1972 par Tulving (Tulving, 1972). Elle demeure fondamentale aujourd'hui. La mémoire épisodique est le système impliqué dans la capacité à se souvenir d'un évènement dans son contexte spatio-temporel d'acquisition, lequel contexte intègre de multiples détails perceptivo-sensoriels et phé-

noménologiques. La mémoire sémantique est le système de mémoire spécialisé dans la connaissance générale du monde (mots, concepts, connaissances factuelles, etc.) indépendamment du contexte d'acquisition.

En 1980, Cohen et Squire ont établi une distinction entre des connaissances déclaratives et des connaissances procédurales (non déclaratives) (Cohen and Squire, 1980). La mémoire déclarative concerne le stockage et la récupération d'informations qu'un individu peut évoquer sous forme d'images et exprimer par le langage. On retrouve dans la mémoire déclarative les connaissances générales (de type sémantique) et spécifiques (de type épisodiques). Récupérer des informations en mémoire déclarative implique un certain niveau de conscience, puisque l'information est amenée à l'esprit et *déclarée*. Au contraire, la mémoire procédurale (aussi appelée mémoire non déclarative) est liée aux savoir-faire (p. ex. *Savoir faire du vélo*). Elle s'exprime par des comportements et actions résultant d'apprentissages passés, et n'implique pas de pensée consciente (p. ex. *Nous faisons du vélo sans conscience des compétences impliquées*).

Tulving s'inspirera des travaux de Cohen et Squire pour amender sa définition des mémoires épisodique et sémantique en intégrant la notion de niveau de conscience associé à la récupération des souvenirs (Wheeler et al., 1997). Une manifestation de la mémoire épisodique, qui permet de revivre un évènement personnellement vécu, est associée à un niveau de conscience autonoétique, qui implique une impression subjective du souvenir et permet la reviviscence consciente de l'évènement à travers un voyage mental dans le souvenir contextualisé (p. ex. *Je me revois en train de déjeuner avec ma mère dans le jardin ; il faisait beau ; j'étais assis au bout de la table en chêne ; elle me servait à boire, puis le dessert*). Par contraste, la manifestation de la mémoire sémantique est associée à un niveau de conscience dit noétique car elle s'accompagne d'une simple conscience de connaissances sur le monde (p. ex. *Je sais qu'Hermann Ebbinghaus est un philosophe allemand, mais je ne me rappelle plus du contexte dans lequel j'ai acquis cette information*).

Dans le cadre de l'étude de la mémorabilité des images en informatique, on pourra s'intéresser à l'épisodicité des souvenirs d'image, à travers leur contextualisation et le niveau de conscience associé à leur récupération (à l'aide des outils présentés dans les sous-sections 1.2.3 et 1.2.3). Ces questions tombent cependant en dehors du champ de celles investiguées dans cette thèse ; nous les aborderons en tant que perspectives pour nos travaux futurs (voir la section 12.5).

Pour conclure sur les différents systèmes de mémoire, bien qu'une taxonomie des systèmes les plus souvent invoqués par les chercheurs pour expliquer les phénomènes de mémoire puisse être proposée (voir la figure 1.2), il faut garder à l'esprit qu'il n'existe pas à ce jour de consensus concernant le nombre ou le fondement des distinctions au sein de la mémoire. De plus, il ne faut pas oublier que, si la mémoire est constituée de plusieurs systèmes de mémoire, ceux-ci fonctionnent en constante interaction : un facteur agissant sur un système de mémoire peut donc avoir une influence sur un autre

système de mémoire, sans que cette influence ne soit forcément perceptible dans les résultats d'une tâche de mémoire.

1.2 Mesurer la mémoire humaine

La mémoire n'est pas un objet directement observable : on ne peut qu'inférer son existence et son fonctionnement de manifestations dont on lui attribue la cause. Des tests et des paradigmes particuliers ont été mis au point pour mesurer spécifiquement les différents types de mémoire.

1.2.1 Les tests de mesure de la performance de la mémoire

Plusieurs tests de mémoire ont été créés pour obtenir une mesure de la mémoire dans des conditions particulières[4]. L'apprentissage du matériel peut être *intentionnel* (on demande explicitement au participant de mémoriser le matériel) ou advenir de manière *incidente* (le participant n'est pas prévenu qu'il va être interrogé sur sa mémoire). La récupération en mémoire peut être *suscitée* (p. ex. demander à un participant de se rappeler un souvenir d'enfance) ou *provoquée* (p. ex. demander à un participant de se rappeler une liste de mots qu'on lui a présentée préalablement). Certaines tâches de mémoire sont dites *directes* ou *explicites*, tandis que les autres sont dites *indirectes* ou *implicites*. Les tâches directes — le rappel libre, le rappel indicé et la reconnaissance — font explicitement référence à un évènement de l'histoire personnelle du participant. Les tâches indirectes — p. ex. la tâche d'identification perceptive, adaptée à l'étude des images — ne font aucune référence explicite à un évènement de l'histoire personnelle du participant (on ne demande pas au participant de rappeler quelque chose), quoiqu'elles soient néanmoins influencées par ces évènements ; elles mesurent un changement de performance (précision, vitesse, etc.) qui est fonction de l'expérience préalable avec les stimuli testés, le type de tâche ou des stimuli reliés aux stimuli testés. Il faut ajouter que les tests de mémoire peuvent être proposés *immédiatement* après l'apprentissage d'un matériel, ou être *différés*— nous y reviendrons dans la prochaine section. En fonction du type de mémoire que l'on désire interroger, on choisira un test plutôt qu'un autre, ainsi que les modalités de son application.

La tâche de **rappel libre** consiste à demander à un participant de rappeler librement quelque chose. Par exemple, on pourra présenter à un participant une série d'images, puis lui demander de rappeler, dans l'ordre qu'il voudra, le thème des images qu'il a vues. Typiquement, la mesure d'un test de rappel libre sera le nombre d'items correctement rappelés sur le nombre total d'items vus (les erreurs ou faux rappels, l'ordre de

[4]Pour une vue détaillée de la question, on pourra lire l'article de (Richardson-Klavehn and Bjork, 1988), sur lequel nous nous sommes appuyés pour écrire cette section.

rappel ou le temps mis pour rappeler les items, peuvent aussi s'avérer intéressants).

La tâche de **rappel indicé** est une tâche de rappel dans laquelle un indice est fourni au participant au moment du rappel. Par exemple, on présente au participant une liste de couples d'images lors d'une phase d'apprentissage — une image cible et une image indice —, puis, lors de la phase de rappel, on présente l'image indice et le participant doit rappeler l'image cible.

La tâche de **reconnaissance** consiste à demander à un participant si un item est apparu précédemment à un moment donné de l'espace et du temps (elle est ainsi différente d'une simple reconnaissance sur la base d'une familiarité). Généralement, dans une première phase dite d'apprentissage, on présente à un participant une liste d'items. Puis, dans une seconde phase dite de récupération, on lui présente ces items vus mélangés à des items non vus, et on va lui demande de reconnaître lesquels il a précédemment vus.

La tâche *d'identification perceptive* est utilisée pour tester la mémoire implicite. L'objectif est de détecter un stimulus présenté sous une forme ou dans des conditions de visualisation dégradées (par exemple, une image dont la qualité aura été dégradée, ou présentée très rapidement). L'idée est que nous reconnaîtrons plus facilement un stimulus qui est bien représenté dans notre mémoire qu'un stimulus qui y est moins bien représenté.

Les scores de mémorabilité des images utilisées pour la mise au point des modèles de prédiction, obtenus par (Isola et al., 2011b), correspondent à une mesure de mémoire réalisée à l'aide d'une tâche de reconnaissance. Cette tâche nous paraît également la plus adaptée à l'étude de la mémorabilité des images en informatique. En effet, pour obtenir des scores sur un nombre d'images conséquent, une tâche de rappel libre ou indicé n'est pas adaptée, d'une part parce qu'il est difficile de décrire verbalement une image, et d'autre part parce que de telles tâches seraient extrêmement difficiles. Quant à la tâche d'identification perceptive, elle mesure une performance de mémoire éloignée des conditions réelles de récupération mnésique des images dans la vie quotidienne (or c'est une mémorabilité généralisable à la vie quotidienne que les chercheurs travaillant sur mémorabilité des images en informatique cherchent à prédire). En outre, des conditions dégradées de visualisation rendraient difficile l'étude des liens entre attention visuelle et mémorabilité, dont on verra qu'elle est intéressante pour la prédiction de la mémorabilité d'image (ce sera l'objet du chapitre 11).

1.2.2 Mesures immédiate et différée

Une des modalité importante d'un test de mémoire est le moment où il est passé. Une performance de mémoire ne sera probablement pas la même selon que la mémoire est interrogée immédiatement après la phase d'apprentissage du matériel, où plus tard. Par exemple, Scapinello et Yarmey ont montré que la performance de reconnaissance d'images était plus basse lorsqu'elle était mesurée 20 minutes après la phase d'appren-

tissage que lorsqu'elle était mesurée immédiatement après (Scapinello and Yarmey, 1970). Les chercheurs dont les travaux portent sur la mémorabilité des images en informatique s'intéressent à une mémorabilité à long terme (du moins, si l'on considère les applications envisagées pour la prédiction de mémorabilité, évoquées dans l'introduction). Comme nous l'avons souligné, la performance de mémoire mesurée par (Isola et al., 2011b), à partir de laquelle sont calculés les scores de mémorabilité des 2222 images constituant leur base de données, a été mesurée quelques minutes après l'encodage mnésique des images : c'est donc une performance de MLT.

Toutefois, les souvenirs stockés en MLT sont encore susceptibles de se modifier. Selon la théorie de la consolidation, il faut un certain temps au cerveau pour stabiliser les traces de mémoire ; durant une phase dite de consolidation à long terme, qui durerait de quelques heures à quelques mois, voire la vie entière, les souvenirs sont fragiles et peuvent être altérés ou modifiés (McGaugh, 2000). Il s'ensuit qu'une image mémorable quelques minutes après son encodage pourrait ne plus être mémorable après un certain temps. D'autre part, si la consolidation ne s'applique pas uniformément à l'ensemble des images (et c'est probablement le cas, comme nous le verrons dans le chapitre 3, où nous expliquons que l'émotion véhiculée par une image pourrait jouer un rôle important dans sa consolidation), il est possible que l'ordre des scores de mémorabilité des images dans la base de données de (Isola et al., 2011b) ne soit plus le même après un certain délai.

Pour s'en rendre compte, il serait intéressant de comparer une performance de mémoire mesurée quelques minutes après la phase d'encodage à une performance de mémoire mesurée après une durée de rétention supérieure à quelques minutes. Pour fixer cette durée de rétention, on pourra s'appuyer sur les courbes d'oubli en MLT. La figure 1.3 représente la courbe d'Ebbinghaus, issue de ses travaux fondateurs sur la mémoire humaine (Ebbinghaus, 1913). On peut remarquer que la perte d'information suit approximativement une courbe exponentielle décroissante : la performance de mémoire (mesurée en quantité de matériel rappelé) diminue très rapidement durant la première heure suivant la phase d'encodage, et dans une moindre mesure durant le premier jour. On pourra, par exemple, choisir de mesurer la mémoire quelques minutes puis un jour après la phase d'encodage : la baisse de mémoire étant importante dans cet intervalle de temps, on pourra se rendre compte si l'ordre de mémorabilité des images en MLT est bouleversé par l'augmentation de la durée de rétention mnésique.

1.2.3 Les techniques complémentaires pour mesurer la mémoire

Des techniques complémentaires ont été développées pour compléter les informations obtenues par les tests standards de mémoire. Nous en présentons trois ici, qui peuvent être utilisées avec une tâche de reconnaissance d'images : le paradigme *what-where-when*, le paradigme *Remember-Know*, et l'échelle de certitude. Les deux premières techniques seraient particulièrement intéressantes dans le cas où on s'intéresserait à

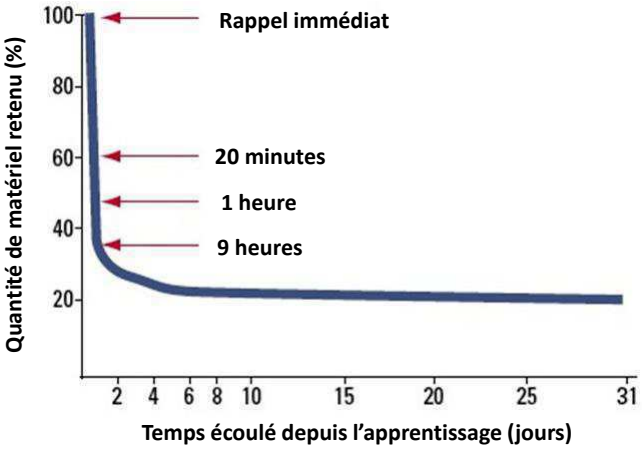


FIGURE 1.3 – La courbe d'oubli d'Ebbinghaus.

l'épisodicité des souvenirs.

Le paradigme *what-where-when*

Dans une optique expérimentale, Tulving a développé un procédé permettant de mesurer les éléments contextuels du souvenir — le Où, le Quand et le Quoi — qui forment le socle du souvenir épisodique et permettent sa remémoration. Ce paradigme, dénommé *what-where-when* en référence aux trois dimensions contextuelles du souvenir, permet de mesurer expérimentalement sa contextualisation. De nombreux travaux ont mis en évidence l'existence d'une mémoire *what-where-when* chez les animaux (voir notamment les travaux de Clayton (Clayton and Dickinson, 1998, Clayton and Dickinson, 1999)) et chez les êtres humains (p. ex. (Plancher et al., 2010). Dans ces derniers travaux, le *what* pouvait être le rappel ou la reconnaissance de l'une ou l'autre des caractéristiques d'un objet présent dans un souvenir (e.g. forme, couleur, etc.) ; le *where* consistait à situer un élément dans son environnement spatial ou par rapport à d'autres éléments ; le *when* était un rappel d'éléments situés dans une chronologie. Les résultats ont montré que l'information *what-where-when* perdurait sur des mois (Plancher et al., 2008) voire des années (Crawley and French, 2005), mais qu'une partie, plus ou moins importante, en était perdue, c'est-à-dire qu'une décontextualisation du souvenir advenait au cours du temps. Dans une tâche de reconnaissance d'images, on pourrait par exemple imaginer demander aux participants, après qu'ils ont reconnu une image, quand celle-ci a été vue durant la phase d'apprentissage, et où elle a été vue sur l'écran (parmi plusieurs localisations possibles).

Il faut noter que la contextualisation du souvenir n'est pas suffisante pour déterminer la nature épisodique du souvenir. Selon Holland et Smulders, la connaissance du *what-where-when* serait une condition nécessaire à la mémoire épisodique mais non pas

suffisante (Holland and Smulders, 2011). Si l'on désire s'assurer de la nature épisodique (ou sémantique) de souvenirs, il sera utile de compléter la mémoire de demander un jugement *remember-know*.

Le paradigme *Remember-Know*

Comme nous l'avons vu, la mémoire épisodique, contrairement à la mémoire sémantique, est associée à une conscience subjective de l'évènement et à la possibilité, à travers elle, de voyager mentalement à travers les souvenirs pour récupérer l'information. Afin de pouvoir déterminer expérimentalement quel type de mémoire — épisodique ou sémantique — un test de mémoire explicite interroge, Tulving, puis Gardiner *et al.*, ont mis au point le paradigme *Remember-Know* (Tulving, 1985, Gardiner et al., 1998). Il permet d'estimer le niveau de conscience associé à une récupération en mémoire.

Le paradigme *Remember-Know* consiste à demander aux participants de qualifier la nature d'une reconnaissance. Ainsi, après avoir reconnu un item comme préalablement vu dans une tâche d'apprentissage, le participant doit dire s'il s'est remémoré l'item, c'est à dire si il est capable de revivre la situation d'apprentissage, d'avoir un souvenir particulier de la présentation de l'item en question. La réponse *Remember* est ainsi clairement associée à un état de conscience autonoétique, et suppose l'existence d'un contexte qui offre un socle à la scène du souvenir ; elle est censée être le reflet d'une utilisation de la mémoire épisodique. Dans le cas contraire, le participant doit répondre qu'il sait que l'item était présent en tâche d'apprentissage, mais sans qu'il puisse récupérer de souvenir particulier de sa présentation ou revivre celle-ci ; on parle alors d'une réponse *Know*, qui serait le reflet de l'utilisation de la mémoire sémantique (Adam, 2003).

Échelle de certitude

Il peut être utile d'interroger un participant sur le degré de certitude associé à ses réponses à un test de mémoire, pour vérifier si le participant est cohérent dans ses réponses. Par exemple, la tâche de reconnaissance de (Isola et al., 2011b), à partir de laquelle ont été obtenus les scores de mémorabilité utilisés pour mettre au point les modèles de prédiction de la mémorabilité d'images, a été passée en crowdsourcing. Dans ces conditions, il est plus difficile de contrôler une étude que dans un laboratoire. Pour cette raison, les auteurs ont mis en place un test de vigilance : certaines images étaient répétées quelques secondes après avoir été vues une première fois, de sorte que, l'information étant toujours en MCT, leur reconnaissance était presque assurée si les participants faisaient preuve de vigilance. En complément de cette technique, on pourrait imaginer utiliser une échelle de certitude, qui permettrait de s'assurer de la cohérence des réponses des participants. Pour ce faire, on pourrait par exemple étudier, pour chaque

participant, les taux de reconnaissance moyens des images associées à chaque degré de certitude, pour s'assurer qu'ils n'ont pas répondu au hasard.

1.3 Conclusion

La mémorabilité des images, telle qu'elle est étudiée en informatique, correspond à une performance de MLT. Cependant, la mesure de la performance de mémoire à partir de laquelle les scores de mémorabilité existants ont été calculés a été réalisée quelques minutes après l'encodage mnésique des images. Or, il n'est pas certain que des images mémorables quelques minutes après leur encodage soient encore mémorables après des délais de rétention plus longs. En effet, la théorie de la consolidation des souvenirs en MLT suggère que la mémorabilité des images pourrait varier dans le temps. Pour investiguer cette question, il serait intéressant de mesurer la MLT à deux moment différents dans le temps ; en particulier, en s'appuyant sur la courbe de l'oubli d'Ebbinghaus, nous avons suggéré qu'en plus d'une mesure réalisée quelques minutes après la phase d'encodage mnésique (similaire à la mesure effectuée par (Isola et al., 2011b)), il serait intéressant d'effectuer une autre mesure un jour après.

La tâche de reconnaissance semble la plus adaptée à l'étude de la mémorabilité des images. Dans une telle tâche, on veillera à présenter les images un temps suffisamment long pour ne pas saturer la MCT, et à intervertir l'ordre des images présentées aux différents participants (notamment, pour éviter l'effet de primauté). On pourra éventuellement utiliser des techniques complémentaires, comme les paradigmes *what-where-when* et *Remember-Know*, pour déterminer si les souvenirs associés aux images sont plutôt épisodiques ou sémantiques, et interroger les participants sur le niveau de certitude associé à leurs réponses pour mieux contrôler l'expérience, en particulier si celle-ci est proposée en crowdsourcing.

Finalement, on privilégiera un raisonnement considérant des performances relatives plutôt qu'absolues, la mémorabilité d'une image n'ayant de sens que par rapport à la mémorabilité d'autres images, mesurée dans des conditions similaires.

Les images sont des vecteurs d'émotion privilégiés. Le chapitre suivant porte sur les émotions humaines, qui jouent un rôle important dans les processus mnésiques.

2

L'image, un vecteur d'émotion

La recherche sur l'émotion est aujourd'hui animée d'un nouveau souffle. L'émergence de l'informatique affective en tant que champ disciplinaire à part entière, dont l'objectif initial fut ainsi posé, « donner aux machines la capacité de reconnaître, d'exprimer, et dans certains cas d'*avoir* des émotions » (Picard and Picard, 1997), suscite un fort engouement de la communauté scientifique. Que l'émotion éveille un tel intérêt chez les chercheurs, particulièrement en informatique (Picard, 2010), n'était pourtant pas évident. Pendant une grande partie de l'histoire de la science et jusqu'à assez récemment, cet objet d'étude a, en effet, largement été tenu à l'écart. Les dernières décennies ont changé la donne : fertiles en études sur l'émotion, elles ont mis en lumière les liens étroits qui unissent l'émotion et la cognition. Il a été montré que l'émotion joue un rôle important dans les processus de décision (p. ex.(Bechara et al., 1999, Bechara et al., 2000)), dans la résolution de problèmes (p. ex. (Trezise and Reeve, 2014)), dans l'attention, dans l'apprentissage et dans les processus mnésiques (p. ex. (Brosch et al., 2013, Christianson, 2014)).

Dans le cadre de l'étude de la mémoire humaine, l'émotion suscitée par un stimulus joue un rôle crucial dans sa mémorisation, à travers plusieurs phénomènes : la sélectivité de l'attention humaine pour les informations émotionnelles (p. ex. (Sharot and Phelps, 2004, Loftus et al., 1987, Fox et al., 2001, Ochsner, 2000) ; la priorisation du traitement des stimuli émotionnels (p. ex. (Kensinger, 2004, Raymond et al., 1992)) ; une meilleure consolidation des souvenirs en MLT pour les stimuli émotionnels (p. ex. (Baddeley, 1982, LaBar and Phelps, 1998)). Les études portant spécifiquement sur la mémorisation d'images confirment l'importance de l'émotion induite par une image dans sa mémorisation (p. ex. (Bradley et al., 1992, Dolcos et al., 2004)).

Pourtant, les études existantes portant sur la mémorabilité des images en informatique ne se sont pas intéressées à l'émotion véhiculée par les images (Isola et al., 2011a, Khosla et al., 2012a, Khosla et al., 2012b, Mancas and Le Meur, 2013, Kim et al., 2013, Celikkale et al., 2013, Oliva et al., 2013, Isola et al., 2014, Bylinskii et al., 2015a, Celikkale et al., 2015, Wang et al., 2015, Lahrache et al., 2016). La question de la répartition dans l'espace émotionnel des images utilisées pour entraîner les modèles prédictifs n'a pas été posée, laissant imaginer de potentiels biais dans les modèles, ou dans les scores de mémorabilité collectés.

La modélisation des liens entre mémoire et émotion revêt un intérêt plus important encore aux yeux de qui cherche à prédire si un matériel va être mémorisé. En effet, les progrès dans notre capacité à mesurer en temps réel les émotions sont rapides, notamment grâce à l'impulsion imprimée à la recherche sur l'émotion en informatique par l'émergence de l'informatique affective. Nous en donnons un exemple dans le chapitre 12, où nous utilisons un dispositif d'électroencéphalographie « grand public » pour mesurer l'émotion de spectateurs d'un film. D'autre part, les chercheurs ont fait quelques progrès dans l'extraction computationnelle de la sémantique émotionnelle des images (Joshi et al., 2011). Ces outils de mesure et d'extraction de l'émotion suscitée ou véhiculée par les images pourraient être mis au service de la prédiction de la mémorabilité des images.

Dans ce chapitre, nous adoptons une définition de l'émotion et présentons les techniques d'induction émotionnelle et de mesure des émotions susceptibles d'être utilisées pour l'étude de l'émotion véhiculée par des images. Nous présentons également, brièvement, le champ de recherche consacré à l'inférence computationnelle de l'information émotionnelle des images.

2.1 Définition de l'émotion

La vérité est dépendante de l'époque. L'histoire des émotions en est une bonne illustration. La notion d'émotion a, en effet, beaucoup varié au cours du temps, depuis les premières théories connues de la Grèce antique et de Chine ancienne, aux théories scientifiques actuelles.

2.1.1 Contexte d'étude de l'émotion

En occident, les théories de l'émotion remontent au moins aussi loin que la Grèce antique. Pour les stoïciens de cette époque, l'émotion était considérée comme un obstacle à la raison et à la vertu. Lorsque l'église deviendra la plus haute instance de décision, au Moyen Âge, la *passion* sera étroitement liée au salut : il faudra désormais vivre les passions qui rapprochent de Dieu et éviter celles qui rapprochent du Diable. Le désenchantement progressif du monde et l'avènement de la pensée mécaniste conduiront à

chercher les causes de l'émotion dans des principes physiques, et non plus seulement métaphysiques.

La première théorie considérée comme scientifique des émotions sera proposée à la fin XIX^{e} siècle par James et Lange, qui défendent au même moment une conception dit « périphéraliste » de l'émotion (Coppin and Sander, 2010). Ces auteurs renversent la cause et la conséquence : ce ne serait plus l'émotion qui déclencherait les manifestations physiologiques, mais la perception de ces manifestations liées au système nerveux périphérique qui déclencherait l'émotion. Dans les années 1930, Cannon et Bard défendront au contraire une conception dite « centraliste » de l'émotion, selon laquelle le déclenchement d'une émotion spécifique est déterminé par le traitement d'un stimulus au niveau du système nerveux central (Coppin and Sander, 2010). Selon Coppin et Sander, les théories James-Lange et Cannon-Bard ont eu un impact considérable sur la considération de l'importance de la cognition dans l'émotion.

A partir des années 1980, le nombre de travaux scientifiques sur les émotions se multiplie. Ces travaux révèleront progressivement le rapport essentiel entre l'émotion et la cognition.

2.1.2 Les trois composantes de l'émotion

Un des problèmes majeurs des sciences sociales est leur utilisation du langage naturel pour la recherche théorique et empirique, qui créé des confusions entre l'acception commune des termes employés et leur acception scientifique. La création artificielle de nouveaux concepts qui ne sont pas contaminés par les connotations du langage naturel est souvent vouée à l'échec, entre autres à cause de la difficulté à obtenir un consensus fort de la part de la communauté scientifique (Scherer, 2005). Ces difficultés sont particulièrement prégnantes dans la recherche sur l'émotion, un terme très utilisé dans la vie quotidienne, pour laquelle les chercheurs peinent à s'accorder sur une définition. Ainsi, un grand nombre de définitions scientifiques de l'émotion ont été proposées (Kleinginna Jr and Kleinginna, 1981).

Dans cette thèse, nous nous en tiendrons à une définition contemporaine de l'émotion, partagée par la plupart des auteurs (Gil, 2009), selon laquelle l'émotion est une réunion complexe de trois composantes. 1/ La composante cognitive, qui correspond aux changements des états mentaux et cognitifs. Cette composante est d'une nature proprement subjective, et ne peut, par conséquent, pas être mesurée directement. 2/ La composante comportementale, qui correspond aux manifestations comportementales, expressives, qui sont particulièrement importantes dans la communication humaine (par exemple, les expressions faciales). 3/ La composante physiologique, qui correspond à l'ensemble des manifestations physiologiques déclenchées en réaction à un stimulus qui suscite une émotion. Le système nerveux autonome, qui assure — sans contrôle de la volonté — l'homéostasie de l'organisme, a un rôle majeur dans les réactions physiologiques liées à l'émotion. En mesurant l'activité du système nerveux autonome, il est

possible d'obtenir une mesure débarrassée des biais induits par l'intermédiaire de la conscience. A noter que ces trois composantes sont liées : elles procèdent d'une même réaction générale d'un individu à un stimulus. Les études peuvent porter sur une, ou plusieurs de ces trois composantes.

2.1.3 Approches catégorielle et dimensionnelle

L'approche dite catégorielle conçoit les émotions comme des entités basiques, discrètes et universelles, dont la liste la plus connue est celle des émotions de base (joie, colère, peur, dégoût, surprise et tristesse) (Ekman, 1992) — liste qu'Ekman élargira en y ajoutant la satisfaction, la honte, l'amusement, le soulagement, etc. (Ekman, 1999). Le recours aux émotions discrètes est fréquent dans la vie courante pour verbaliser son ressenti ; il est aussi courant dans la littérature scientifique. Cependant, pour certains auteurs, établir des catégories émotionnelles subjectives pourrait biaiser notre vision de ce qu'est réellement l'émotion (Barrett, 2006).

L'approche dimensionnelle repose sur l'idée que les émotions peuvent être appréhendées par des dimensions élémentaires, qui correspondraient à des propriétés phénoménologiques basiques de l'expérience affective (Russell and Barrett, 1999). Deux dimensions émotionnelles sont beaucoup plus étudiées que les autres : l'arousal et la valence[1] Elles permettent de différencier la plupart des émotions discrètes classiquement étudiées (Posner et al., 2005). L'arousal se définit généralement sur un continuum calme-excitation (Lang et al., 1997, Bradley and Lang, 1994), et fait référence au degré d'éveil du sujet. Cette dimension est parfois appelée *activation* (p. ex. (Gil, 2009)) ; on parlera alors de stimuli activateurs pour évoquer les stimuli suscitant de l'arousal. La valence se définit sur un continuum déplaisir-plaisir (négatif-positif), et correspond au degré de satisfaction et de bien-être du sujet (Revelle and Loftus, 1992). Étant donnée la difficulté à identifier une troisième dimensions (par exemple, de contrôle ou de tension) distinguable de l'arousal, la plupart des théoriciens modernes de l'approche dimensionnelle se cantonnent aux dimensions d'arousal et de valence (Scherer, 2005).

Puisque les émotions discrètes peuvent être appréhendées par des dimensions émotionnelles, certains auteurs ont proposé de faire coïncider les émotions discrètes avec des scores émotionnels sur les dimensions d'arousal et de valence (p. ex. (Posner et al., 2005) ; voir la figure 2.1).

Dans cette thèse, nous adoptons une approche dimensionnelle des émotions. Ce choix repose sur plusieurs arguments, évoqués par (Scherer, 2005).

D'abord, l'approche dimensionnelle est généralement fiable (Scherer, 2005). Les scores d'arousal et de valence obtenus dans différentes études pour les mêmes images

[1] Il n' a pas de consensus sur la traduction de ces termes en français. L'arousal est parfois traduit par « activation », « excitation », ou encore « éveil ». Le mot valence n'a, quant à lui, pas d'équivalent en français. Pour cette raison, nous préférons conserver la dénomination anglaise.

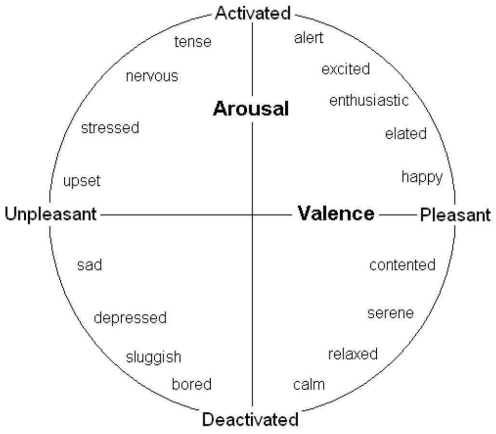


FIGURE 2.1 – Représentation graphique du circomplexe arousal (axe vertical)-valence (axe horizontal). Les émotions discrètes sont positionnées sur la base de leur score dans ces deux dimensions. (Tirée de (Posner et al., 2005).)

montrent de bonnes corrélations (particulièrement pour la dimension de valence). Nous détaillerons ce point dans le chapitre 6, où nous comparons les scores d'arousal et de valence obtenus dans notre étude pour 150 images aux scores obtenus pour les mêmes (ou partie des mêmes) images obtenus dans plusieurs études précédentes (Lang et al., 1997, Ito et al., 1998, Lang et al., 2008, Grühn and Scheibe, 2008). La comparaison inter-études est généralement plus difficile pour les études qui s'inscrivent dans une approche catégorielle, notamment à cause de la grande variété existant dans le nombre et le choix des émotions discrètes mesurées (Scherer, 2005), ces modalités pouvant avoir une influence importante sur les mesures de l'émotion obtenues.

De plus, toujours selon Scherer, par rapport à l'approche catégorielle, l'approche dimensionnelle permet un traitement statistique plus poussé des résultats obtenus par des questionnaires d'auto-évaluation. La raison en est qu'il est plus simple, dans cette dernière approche, d'utiliser des échelles graduées dont les intervalles sont réguliers (voir par exemple les échelles 2.4 dans la section 2.3.1). L'adoption d'une approche dimensionnelle plutôt qu'une approche catégorielle est donc susceptible de faciliter l'analyse conjointe des scores de mémorabilité et d'émotion d'images (point que nous aborderons dans le chapitre 7).

Enfin, nous sommes particulièrement intéressés par les mesures physiologiques (pour les raisons avancées dans la partie 2.3.2). Or, ces mesures sont beaucoup plus adaptées pour mesurer l'arousal que des émotions discrètes. C'est peut-être d'ailleurs la raison principale de l'engouement des chercheur en informatique affective — qui font une large utilisation des mesures physiologiques pour mesurer en temps réel les émotions d'individus — pour cette dimension émotionnelle (et, partant, pour l'approche dimensionnelle). Dans le chapitre 12, nous utilisons un dispositif d'électroencéphalographie

(Emotiv,) qui fournit directement des données d'arousal.

Il est également important de souligner les deux défauts majeurs de cette approche.

D'une part, lorsqu'il est demandé à un individu d'évaluer la valence de l'émotion induite par un stimulus — une image, par exemple — à l'aide d'une échelle, il est difficile de savoir s'il a vraiment mesuré son état émotionnel, ou plutôt la qualité intrinsèque du stimulus. Par exemple, si on présente une image de serpent à un individu, et qu'il attribue une note négative à l'émotion suscitée par cette image, on ne pourra pas être certain que cette note porte exclusivement sur l'émotion suscitée par l'image, ou si elle se rapporte plutôt à sa représentation mentale du serpent. Or, la valence comme qualité intrinsèque d'un stimulus et la valence comme qualité de l'émotion ressentie par l'individu ne coïncident pas nécessairement (Scherer, 2005). Pour diminuer cette incertitude, la consigne d'une tâche de notation devra être claire sur ce point, et les outils utilisés devront favoriser une bonne compréhension de ce qui est attendu par les expérimentateurs. Dans notre expérience présentée dans le chapitre 5, nous utilisons des échelles graphiques (voir la figure 2.4 dans la section 2.3.1) qui présentent de petits dessins qui symbolisent des hommes sujets à des émotions de degré variable, et rappellent que la notation doit porter sur le ressenti émotionnel. Malgré de telles précautions, les interprétation relatives à ce point particulier devront être faites avec prudence. On notera qu'il pourrait être intéressant de croiser une mesure de la composante cognitive avec une mesure de la composante physiologique (voir la section 2.3.2) ou comportementale (voir la section 2.3.2), qui dépendent moins de l'interprétation des individus.

D'autre part, contrairement à l'approche catégorielle où les termes employés pour l'auto-évaluation (p. ex. joie, colère, tristesse) sont aisément compréhensibles par quiconque, les dimensions d'arousal et de valence peuvent être assez ambiguës. La consigne devra, par conséquent, expliciter clairement ce que représentent l'arousal et la valence. Il sera également important de familiariser les participants avec la tâche de notation, et de s'assurer, à la fin de cette phase d'entraînement, que le participant a bien compris ces notions. Les échelles graphiques que nous utilisons dans l'expérience présentée dans le chapitre 5, permettent également de faciliter la compréhension de ce qui est attendu, par rapport à des échelles verbales.

Relations géométriques entre l'arousal et la valence

Si la grande majorité des études convergent vers la même conclusion, que l'arousal et la valence sont fondamentales pour la nature de l'affect (Carroll et al., 1999, Kuppens et al., 2013), la relation entre ces deux dimensions émotionnelles n'est pas toujours évidente. Ainsi, plusieurs modèles théoriques supposent que l'arousal et la valence sont des dimensions indépendantes qu'ils décrivent géométriquement comme des dimensions orthogonales (notamment (Barrett and Russell, 1999,Carver and Scheier, 1990,Larsen and Diener, 1992) ; voir la figure 2.2(a)). Pour quelques auteurs, l'arousal covarie positive-

ment avec la valence (p. ex. (Pettinelli, 2008) ; voir la figure 2.2(b)) ou, au contraire, négativement (p. ex. (Tsai et al., 2006) ; voir la figure 2.2(c)). D'autres auteurs pensent que l'arousal et la valence sont liés par une relation en forme de V ou de U symétrique (p. ex. (Jennings et al., 2000, Bernat et al., 2006, Bradley et al., 2001) ; voir la figure 2.2(d)) ou asymétrique, très souvent avec les images négatives associées à un arousal plus fort que les images positives (p. ex. (Ito and Cacioppo, 2005, Ito et al., 1998, Baumeister et al., 2001) ; voir la figure 2.2(e)).

La relation géométrique entre l'arousal et la valence dépend notamment du type de stimuli utilisés pour induire l'émotion (Kuppens et al., 2013). Pour des images, la plupart des auteurs trouvent une relation en U ou V entre l'arousal et la valence (Bernat et al., 2006, Bradley et al., 2001, Ito et al., 1998, Libkuman et al., 2007, Bradley and Lang, 2007, Lang, 1995, Lang et al., 1999, Lang et al., 2008). Il existe quelques exceptions cependant (p. ex. (Ribeiro et al., 2005, Grühn and Scheibe, 2008), avec des auteurs qui trouvent une relation négative plutôt linéaire entre l'arousal et la valence : les images négative sont évaluées comme suscitant plus d'arousal que les images neutres, alors que les images positives ne diffèrent pas, voir suscitent moins d'arousal que les images neutres [2]. Nous participerons à cette discussion sur la base de nos résultats présentés dans la section 6.2 du chapitre 6.

2.2 Induction d'émotions

Afin d'étudier les émotions dans une situation contrôlée, c'est-à-dire qui leur permet de manipuler ou de fixer les différentes variables d'influence de leurs études, les psychologues ont développé un certain nombre de méthodes leur permettant de susciter des émotions particulières chez des individus (Kučera and Haviger, 2012). On appelle techniques d'induction émotionnelle ces méthodes destinées à la manipulation contrôlée des émotions d'individus.

2.2.1 Techniques d'induction émotionnelle

Les techniques d'induction émotionnelle sont nombreuses. Les techniques dites standardisées reposent sur l'utilisation d'un matériel d'induction évalué auprès d'un échantillon représentatif d'individus. Cette évaluation a pu être réalisée à l'aide d'un ou plusieurs types de mesure (par exemple, les émotions véhiculées par les images de l'International Affective Picture System (Lang et al., 2008) ont été évaluées à l'aide de mesures des composantes cognitive, physiologique et comportementale de l'émotion, comme nous l'expliquerons dans la section 2.2.2). Il existe des bases de données standardisées d'images (p. ex. (Lang et al., 2008)), de vidéos (p. ex. (Gross and Levenson,

[2] Dans ce cas, la notation a été réalisée par des adultes âgés (voir (Grühn and Scheibe, 2008))

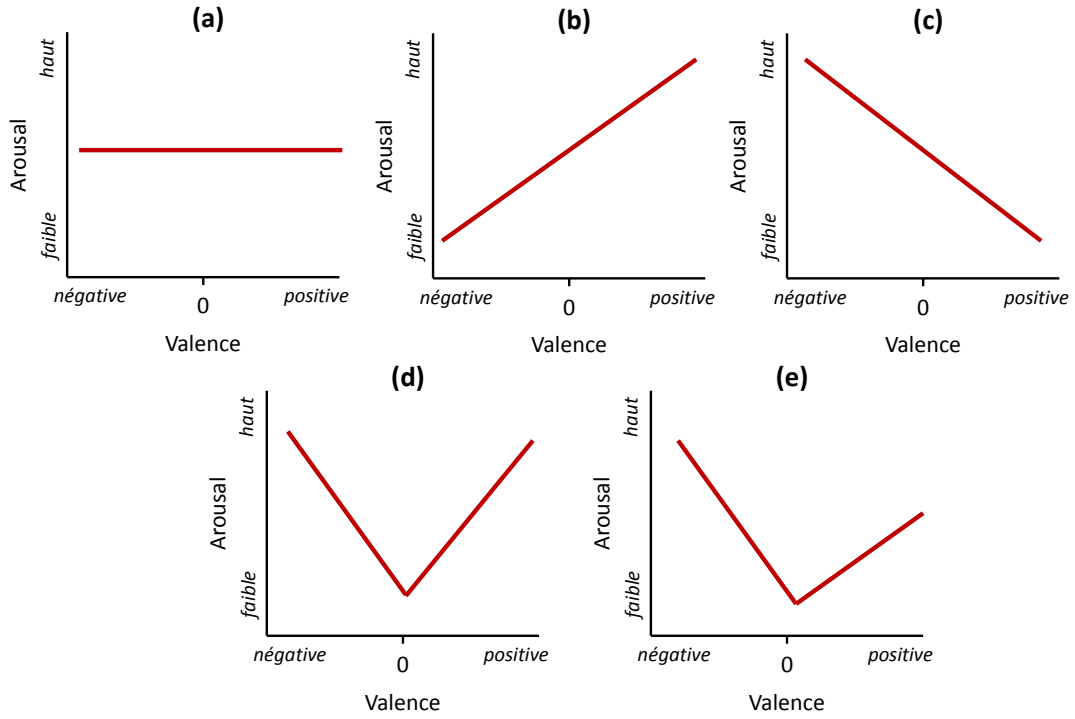


FIGURE 2.2 – Vue d'ensemble des relations trouvées entre Arousal et Valence : (a) indépendance, (b) relation linéaire positive, (c) relation linéaire négative, (d) relation en V symétrique, et (e) relation en V asymétrique avec un biais négatif (inspiré de (Kuppens et al., 2013)).

1995, Schaefer et al., 2005)), et, depuis plus récemment, d'extraits musicaux (Vieillard et al., 2008), utilisées pour induire des émotions spécifiques. D'autres techniques sont parfois utilisées ; par exemple, le rappel autobiographique (i.e. demander au participant de se remémorer le souvenir d'un événement associé à une émotion particulière (Smith and Ellsworth, 1985)), ou la manipulation du succès ou de l'échec (i.e. rendre compte au participant de son résultat à une tâche qu'il vient d'effectuer et pour laquelle il ne peut évaluer sa propre performance, en déclarant ce résultat satisfaisant ou, au contraire, décevant ; p. ex. (Shapiro and Herbert, 1967)).

En accord avec l'esprit de concision qui anime l'écriture de cette première partie de thèse, nous nous en tiendrons à la présentation des bases de données standardisées d'images, spécialement conçues pour l'induction émotionnelle. On notera cependant que des techniques d'induction émotionnelle qui ne reposent pas sur l'utilisation d'images pourraient également trouver leur place dans le cadre de l'étude de la mémorabilité d'images ; par exemple, pour évaluer si la mémorisation d'une image neutre est facilitée par l'induction d'une émotion induite par un autre élément de l'environnement durant sa présentation (par exemple, en présentant une musique émotionnellement colorée pendant la visualisation de l'image), ou encore pour évaluer la compétition entre différents sens (par exemple, auditif et visuel) dans un cadre de mémorisation d'images.

2.2.2 Bases de données d'images standardisées

Trois bases de données standardisées d'images sont, à notre connaissance, disponibles à ce jour : l'International Affective Picture System (IAPS), constituée de 1196 images (Lang et al., 1997,Lang et al., 1999,Lang et al., 2008), la Geneva Affective Pictures database (GAPED), constituée de 730 images (Dan-Glauser and Scherer, 2011), et la Nencki Affective Picture System (NAPS), qui comprend 1356 images (Marchewka et al., 2014).

L'IAPS est la base de données d'images la plus ancienne, et, de très loin, la plus utilisée. Elle est composée de photographies qui présentent des contenus variés (végétation, accidents, cadavres, nourriture, animaux, objets inanimés, pornographie, etc.). Chaque image a été initialement évaluée sur les dimensions émotionnelles d'arousal, de valence et de dominance [3] au moyen des échelles SAM (pour *Self-Assessment Manikin* ; voir la figure 2.4 dans la section 2.3.1).

Plusieurs chercheurs ayant utilisé tout ou partie des images de l'IAPS ont rendu disponibles les résultats de leurs études, augmentant cette base de données d'informations intéressantes. En particulier, il est possible de télécharger des scores d'émotions discrètes (Mikels et al., 2005, Libkuman et al., 2007) et de mémorabilité (Libkuman et al.,

[3] La dominance, parfois appelée contrôle, qui renvoie à la sensation du sujet de pouvoir ou non contrôler la situation, est une dimension émotionnelle secondaire, moins fortement liée à l'état émotionnel que l'arousal et la valence (Lang et al., 2008)

2007, Grühn and Scheibe, 2008) pour certaines images de l'IAPS. Pour ce qui est de ces scores de mémorabilité, ils sont, à notre connaissance, les seuls disponibles pour des images en dehors de la base de données de (Isola et al., 2011b), et les seuls disponibles pour des images évaluées en matière d'émotion. Les scores de mémorabilité proposés par Grühn et Scheibe ont été obtenus grâce à une tâche de reconnaissance ; la méthodologie adoptée (détaillée dans (Grühn et al., 2007)) est cependant assez spécifique à leur question de recherche, et s'éloigne de la méthode employée par (Isola et al., 2011b). Nous reviendrons plus en détails sur ce point dans la section 7.1 du chapitre 7. Les scores de mémorabilité proposés par Libkuman *et al.* correspondent à des jugements portés a priori sur le degré de mémorabilité d'une image. Ces auteurs ont demandé aux participants à leur étude d'évaluer à l'aide d'une échelle à neuf degrés à quel point des images qu'ils voyaient pour la première fois leurs paraissaient mémorables. Les scores de mémorabilité calculés à partir des résultats ne correspondent donc pas à une performance de mémoire. Ils présentent cependant un intérêt dans le cadre de l'étude de la mémorabilité des images en informatiques, où ils pourraient nous donner une idée de la capacité des individus à prédire le degré réel de mémorabilité d'une image — donc, indirectement, de la précision de l'annotation manuelle des images en matière de mémorabilité. Nous reviendrons également sur ce point dans la section 7.1 du chapitre 7.

Contrairement aux autres bases de données précitées, tout ou partie des images de l'IAPS ont été réévaluées sur les dimensions d'arousal et de valence (p. ex. (Grühn and Scheibe, 2008, Ito et al., 1998, Libkuman et al., 2007, Ribeiro et al., 2005, Soares et al., 2014)). Les résultats obtenus dans ces différentes études sont cohérents, ce qui signifie que, d'une manière générale, les mêmes images suscitent les mêmes émotions. En outre, certaines des images de l'IAPS ont été évaluées sur les dimensions d'arousal et de valence par différents groupes d'individus (p. ex. adultes jeunes et âgés (Grühn and Scheibe, 2008), personnes de nationalités différentes (Soares et al., 2014)), à différentes époques (p. ex. dans les années 1990 par (Ito et al., 1998), dans les années 2000 par (Libkuman et al., 2007) et (Grühn and Scheibe, 2008), dans les années 2010 par (Soares et al., 2014)). Les effets de l'induction émotionnelle par les images de l'IAPS ont également été évalués par des mesures de l'activité physiologique (p. ex. mesure de la conductance de la peau, du rythme cardiaque, de l'intensité du sursaut (Bradley et al., 2001)), neurophysiologique (p. ex. mesure des potentiels évoqués [4], au moyen de l'électroencéphalographie (EEG) (Cuthbert et al., 2000)), par des études en IRMf (p. ex. (Lang et al., 1998)), ou encore par des mesures de la modulation de la taille de la pupille (p. ex. (Bradley et al., 2008)).

Pour ces raisons, l'IAPS apparaît comme une solution satisfaisante pour étudier les liens entre la mémorabilité d'images et l'émotion qu'elles véhiculent.

[4] Les potentiels évoqués, ou ERP — pour *Event Related Potentials* —, correspondent à la modification de l'activité électrique produite par le système nerveux en réponse à une stimulation externe

L'inventaire dressé des qualités de l'IAPS ne doit cependant pas occulter ses faiblesses. On pourrait notamment arguer, pour ne pas la choisir, du manque de naturel ou de la qualité toute relative de certaines images et du manque d'actualité des scènes représentées dans quelques-unes des images ; sur ces points, la GAPED et la NAPS, plus récentes, font beaucoup mieux. On pourrait ajouter — mais il en est de même pour les deux autres bases — que l'IAPS est une base de données assez petite au regard de celles que l'on peut trouver pour l'apprentissage machine. La variété de thèmes est également assez limitée (par exemple, il n'est pas rare de tomber sur un serpent en visionnant les images).

2.3 Mesure des émotions

Mesurer l'état émotionnel d'une personne est un des problèmes les plus épineux des sciences affectives modernes (Mauss and Robinson, 2009). Dans le cadre de notre définition suivant laquelle l'émotion est une réunion complexe de trois composantes — cognitive, physiologique et comportementale —, les mesures de l'émotion peuvent être distinguées selon qu'elles portent sur l'une ou l'autre de ces composantes. Les mesures de la composante cognitive sont subjectives : elles reposent sur le jugement qu'un individu fait de son propre état émotionnel. Typiquement, la composante cognitive est évaluée à l'aide d'instruments d'auto-évaluation (p. ex. (Mayer and Gaschke, 1988, Izard, 1993, Bradley and Lang, 1994, Mehrabian, 1996). Les mesures de la composante physiologique sont, au contraire, objectives. Elles sont réalisées à l'aide de capteurs, tels que l'EEG (Horlings et al., 2008) ou le galvanomètre (pour mesurer l'activité électrodermale) (Bradley et al., 2001). Quant aux mesures de la composante comportementale, moins utilisées que les précédentes, elles correspondent généralement à des mesures des expressions faciales (Ekman and Friesen, 1977)). La plupart des méthodes de mesure de l'état émotionnel se concentrent sur une de ces trois composantes de l'émotion. Il est cependant important de garder à l'esprit que réponses cognitive, physiologique et comportementale procèdent d'une même réaction générale d'un individu à un stimulus vecteur d'émotion. On ne mesure donc généralement qu'une *partie de l'éléphant* (voir la figure 2.3).

2.3.1 Questionnaires et échelles d'auto-évaluation

Plusieurs questionnaires, ont été créés pour permettre à des individus d'évaluer leur ressenti émotionnel (à l'aide d'échelles d'auto-évaluation) en les dirigeant (à l'aide des questions) sur les composantes de celui-ci.

Certains questionnaires mesurent directement des émotions discrètes. Par exemple, la DES (pour *Differential Emotions Scale*) est composées de trente adjectifs qui correspondent à dix états émotionnels (joie, tristesse, peur, colère, etc.), et l'évaluation se fait

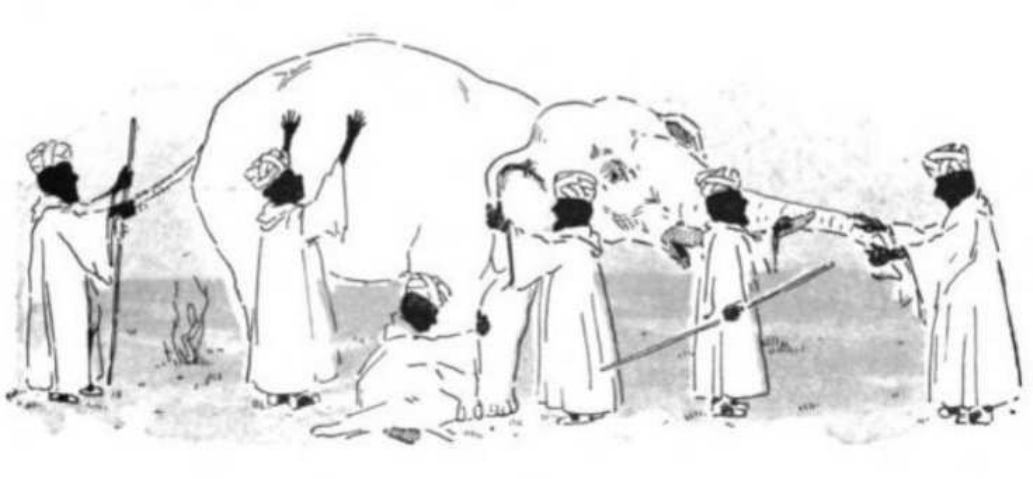


FIGURE 2.3 – Des hommes aveugles et un éléphant. Tiré de (Stebbins and Coolidge, 1909).

sur une échelle à cinq degrés (Izard, 1993). Autre exemple, la BMIS (pour *Brief Mood Introspection Scale*), qui est composée de seize adjectifs et d'une échelle à quatre degrés. Elle permet d'obtenir, selon le calcul effectué, des scores émotionnels discrets ou dimensionnels (Mayer and Gaschke, 1988).

D'autres questionnaires se concentrent uniquement sur les dimensions émotionnelles. Ainsi en est-il de la PAD (pour *Pleasure-Arousal-Dominance*), qui est composée de dix-huit phrases — six pour chacune des dimensions suivantes : la valence, l'arousal et la dominance — pour chacune desquelles le sujet doit juger, à l'aide d'une échelle en sept points, de son degré de correspondance avec son propre état émotionnel (Mehrabian, 1996).

La technique d'auto-évaluation de l'état émotionnel la plus populaire (au regard du nombre de citations des articles présentant les différentes techniques de mesure présentées ici) est la SAM (pour *Self-Assessment Manikin*) (Bradley and Lang, 1994). Elle comprend plusieurs échelles graphiques — pour la valence, l'arousal et la dominance [5] —, dont chacune correspond à la discrétisation en neuf degrés sur la dimension considérée de l'état émotionnel d'un individu schématiquement représenté (voir la figure 2.4). La SAM est particulièrement intéressante. En effet, elle répond aux principales critiques faites aux échelles verbales : manque de pertinence de certains items verbaux ; biais inhérents à la compréhension du vocabulaire ; grande difficulté à traduire fidèlement ces échelles, ce qui pose notamment un problème pour les comparaisons inter-culturelles ; difficulté pour utiliser correctement ces échelles avec des personnes — en particulier les

[5] La dimension de dominance, ou contrôle, renvoie à la sensation du sujet de pouvoir contrôler la situation. Comme nous l'avons précédemment évoqué, cette dimension est souvent laissée de côté par les théoriciens modernes de l'approche dimensionnelle (Scherer, 2005).

enfants — dont le lexique émotionnel ne permet pas une compréhension satisfaisante de l'ensemble des items verbaux (Gil, 2009). Dans le cadre d'une étude proposée en crowdsourcing, à l'instar de celle de (Isola et al., 2011b), où les participants ont généralement des origines très variées, la SAM sera particulièrement intéressante.

Les techniques d'auto-évaluation sont généralement critiquées sur les aspects suivants : d'abord, nous ne sommes pas tous égaux pour comprendre et verbaliser nos émotions ; d'autre part, certaines personnes présentent une tendance à répondre ce qu'on attend d'elles (on parle, dans ce cas, de biais d'attente) ; enfin, l'évaluation nécessite de porter sa conscience sur son ressenti, ce qui peut modifier ce dernier. Dans certains cas, il pourra être plus intéressant d'utiliser, par exemple, une mesure de la composante physiologique de l'émotion, moins sujette à ces biais (ou d'avoir recours à une combinaison de mesures).

Ces outils sont cependant largement utilisés par les chercheurs (l'article présentant la SAM (Bradley and Lang, 1994), par exemple, a été cité près de 3800 fois, d'après les chiffres de Google Scholar).

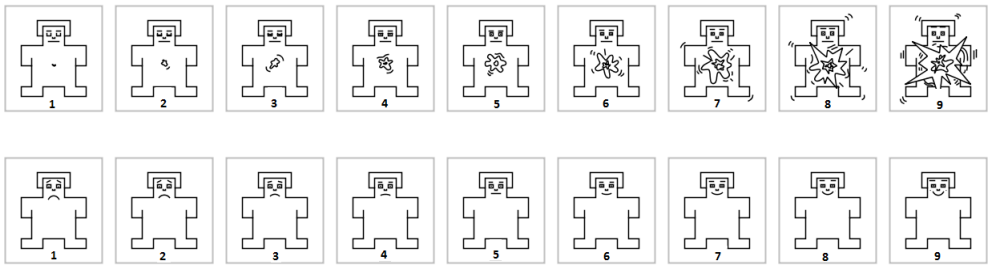


FIGURE 2.4 – Les échelles graphiques SAM pour l'évaluation des dimensions émotionnelles d'arousal (en haut) et de valence (en bas) (Bradley and Lang, 1994).

2.3.2 Mesures des composantes physiologique et comportementale de l'émotion

Un des points d'intérêt de l'informatique affective est d'aider les hommes à mieux comprendre leurs propres émotions en utilisant des outils informatiques (Picard, 2010), car les hommes ont parfois des difficultés à comprendre — et a fortiori à évaluer — leurs propres émotions. L'intérêt des chercheurs en informatique affective pour les mesures physiologiques vient en partie de cette réalité. Ces mesures permettent en effet, dans certaines situations, d'obtenir des informations sur l'état émotionnel de personnes, qu'elles-mêmes ne seraient pas en mesure de mettre en conscience (Picard, 2010). D'autre part, par rapport à des mesures réalisées à l'aide d'outils d'auto-évaluation, les mesures de la composante physiologique de l'émotion sont moins invasives (Gil,

2009), et elles permettent de mesurer l'émotion en temps réel. Ce dernier point rend particulièrement intéressantes les mesures physiologiques dans le cadre de l'étude de la mémorabilité en informatique, pour étendre les travaux sur les images aux vidéos (nous reviendrons sur ce point dans le chapitre 12).

Ces techniques présentent toutefois plusieurs faiblesses. Premièrement, les patrons de réponses physiologiques correspondant aux différents états émotionnels sont encore mal définis. Ce point évolue rapidement, cependant, grâce au nombre important d'études menées portant sur ce type de mesure (Silva et al., 2015). D'autre part, le protocole expérimental inhérent aux techniques de mesure physiologiques peut provoquer des réactions émotionnelles parasites (p. ex. stress face à l'appareillage). Par ailleurs, l'activité physiologique mesurée par ces techniques n'est pas uniquement due à la réponse émotionnelle des individus, mais également à d'autres fonctions de l'organisme liées à la digestion, l'homéostasie, l'effort, l'attention, etc. (Berntson and Cacioppo, 2000). Enfin, si les réponses physiologiques, qui donnent des informations sur l'état émotionnel relativement à son arousal et/ou sa valence, sont bien adaptées à une approche dimensionnelle des émotions, elles sont moins adaptées à l'étude des émotions discrètes (Gil, 2009).

Les principales mesures de la composante physiologique de l'émotion portent sur : les potentiels évoqués, la réponse électrodermale, la dilatation pupillaire, la fréquence cardiaque, la fréquence respiratoire, les variations de la température corporelle, la concentration de certaines hormones ou neurotransmetteurs, la pression sanguine, l'asymétrie de l'activité corticale. La liste n'est pas exhaustive ; nous ne décrirons plus en détails que les techniques les plus couramment utilisées.

Les techniques de mesure de la composante comportementale de l'émotion sont plus marginalement utilisées. Elles reposent principalement sur la mesure des expressions faciales (Ekman and Friesen, 1977), possiblement à l'aide d'un électromyogramme (Witvliet and Vrana, 1996). La mise en œuvre de ces techniques, et l'interprétation des mesures, sont complexes.

L'électroencéphalographie

L'EEG est une méthode d'enregistrement de l'activité électrique du cerveau au moyen d'électrodes placées sur le scalp. L'analyse des signaux électriques recueillis est susceptible de nous renseigner sur les processus cérébraux sous-jacents à une réaction à la présentation d'un stimulus. Cette technique a été utilisée avec un certain succès pour évaluer l'état émotionnel de personnes à la fois en matière d'émotions discrètes (Li et al., 2009) et de dimensions émotionnelles (Horlings et al., 2008).

Depuis quelques années, plusieurs dispositifs EEG « grand public » ont été mis sur le marché (Cernea et al., 2011). Leurs prix sont relativement abordables et ils sont simples d'utilisation. C'est le cas du casque Emotiv Epoc, que nous utilisons dans le film interactif « émotionnel » présenté dans le chapitre 12, à l'origine destiné au monde du jeu

vidéo et conçu comme une interface cerveau-machine (Emotiv,). Le logiciel d'Emotiv associé à ce dispositif fournit directement, en plus des données brutes, des valeurs sur plusieurs dimensions émotionnelles (i.e. l'excitation, la frustration, la méditation et l'engagement). Si ce type de casque était amené à se répandre dans les foyers, il pourrait notamment être utilisé pour annoter automatiquement un grand nombre d'images numériques (i.e. leur attacher des métadonnées relatives à l'émotion qu'elles véhiculent). La fiabilité de ces alternatives grand public n'est cependant pas encore bien établie, même si plusieurs études ont montré qu'ils étaient — au moins passablement — fonctionnels (Duvinage et al., 2013, Ekanayake, 2010, Stytsenko et al., 2011).

Oculométrie et variations pupillaires

L'oculométrie regroupe un ensemble de techniques qui permettent d'enregistrer les mouvements oculaires. Il est possible d'inférer de tels enregistrements les positions successives du regard porté sur une image et, partant, d'obtenir une mesure du déploiement de l'attention visuelle dans l'image. Or, les propriétés émotionnelles des images jouent un rôle important dans l'exploration visuelle et la répartition de l'attention (Quirk and Strauss, 2001). Elles peuvent aider à organiser l'exploration visuelle, en dirigeant l'attention vers les régions les plus informatives (Christianson et al., 1991). De manière plus globale, les images suscitant de l'arousal sont plus largement explorées que les images neutres (Quirk and Strauss, 2001), et les images suscitant de l'émotion occasionneraient également un nombre de fixations plus important que les images neutres (Carniglia et al., 2012). D'autre part, une étude récente (Mancas and Le Meur, 2013) suggère que l'attention visuelle portée à une image est liée à sa mémorabilité. Nous reviendrons en détail sur le lien entre attention visuelle, mémorabilité et émotion dans la section 11.2 du chapitre 11.

D'autre part, les appareils qui permettent de mesurer les mouvements des yeux, appelés oculomètres, fournissent généralement une mesure dynamique de la dilatation pupillaire. Or, la dilatation pupillaire est un indice de l'état émotionnel d'un individu. Elle est liée à la fois aux variations d'arousal (Bradley et al., 2008) et de valence (Partala and Surakka, 2003).

Ces deux types de mesure, non invasives, font de l'oculomètre un outil de choix pour étudier l'émotion suscitée par des images, en lien avec leur mémorabilité.

La réponse électrodermale

La réponse électrodermale (ou réflexe psychogalvanique) correspond aux changements des propriétés électriques de la peau humaine en réponse à une stimulation. L'activité des glandes sudoripares eccrines (nombreuses dans les mains, elles sont responsables de la sudation) s'accompagne d'une modification de la conductance à la surface de la peau, qui peut être mesurée avec un capteur spécifique appelé galvanomètre. Or, cette

activité est liée à l'état émotionnel. Ce résultat a, par exemple, été montré pour des images de l'IAPS (Bradley et al., 2001). Dans cette étude, Bradley *et al.* montrent que les images de valences positive et négative occasionnent des réponses électrodermales plus fortes que les images neutres. La figure 2.5 représente un enregistrement d'une réponse électrodermale pendant 16 secondes, à partir du début de la présentation soit d'un mot neutre, soit d'un juron. Cet enregistrement a été effectué par (Bowers and Pleydell-Pearce, 2011).

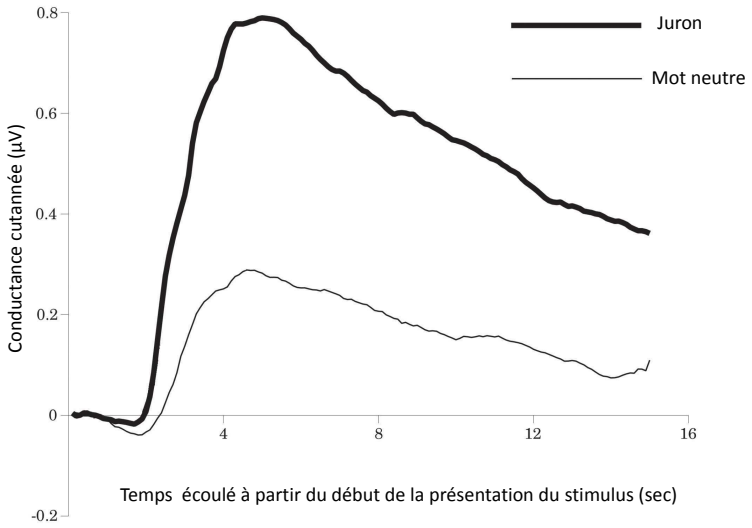


FIGURE 2.5 – Exemple d'enregistrement d'une réponse électrodermale sur une durée 16 secondes. L'ordre de grandeur du potentiel électrique mesuré par le galvanomètre sur l'axe des ordonnées est le millivolt. Les résultats ont été tirés de (Bowers and Pleydell-Pearce, 2011).

Les réponses cardiaque et respiratoire

La réponse cardiaque à une stimulation est typiquement mesurée à l'aide d'un électrocardiogramme. Le rythme cardiaque est particulièrement informatif dans l'étude de l'émotion (nombre de battements par minutes, accélérations/décélérations) ; en effet, il a depuis longtemps été montré qu'un changement d'état émotionnel s'accompagnait de variations dans le signal mesuré par l'électrocardiogramme (Stevenson et al., 1950).

La mesure de la fréquence respiratoire est généralement obtenue au moyen d'une ceinture thoracique. Elle correspond au nombre de soulèvements du thorax pendant une minute (nombre de cycles respiratoires — inspirations et expirations — par minute). Certaines émotions discrètes induisent des changements spécifiques de la fréquence

respiratoire (Homma and Masaoka, 2008). Une respiration plus rapide a également été observée suite à une induction émotionnelle par des stimuli véhiculant un arousal fort (Nyklíček et al., 1997, Boiten, 1998).

La mesure des expressions faciales

Notre faculté à reconnaître un visage et l'émotion qu'il véhicule joue un rôle essentiel dans nos interactions sociales (Chaby and Narme, 2009). D'un point de vue évolutionniste, cela pourrait expliquer la force avec laquelle un visage peut exprimer des émotions. Les expressions faciales, en particulier, sont un puissant vecteur d'émotions. L'étude de la composante comportementale de l'émotion repose généralement sur la mesure des expressions faciales. Les deux techniques de mesure les plus répandues sont le *Facial Action Coding System* (FACS) et l'électromyographie de surface (EMG de surface).

Le FACS repose sur l'idée que chaque émotion correspond à un schéma d'activation musculaire qui lui est propre (Ekman and Friesen, 1977). Les expérimentateurs formés à son utilisation codent, à partir d'images ou de vidéos, la contraction ou la détente des muscles faciaux sur la base d'unités d'action (UA ; voir la figure 2.6). Les mouvements du visage sont décomposés en 46 UA ; l'expression d'une émotion est composée d'un ensemble d'UA. Par exemple, l'UA 6 correspond à la remontée des joues due à la contraction du muscle orbiculaire de l'œil, l'UA 12 correspond à l'étirement du coin des lèvres dû à la contraction du muscle grand zygomatique ; la joie correspond à l'UA 6 plus l'UA 12.

Dans le cadre de la mesure des expressions faciales, l'électromyographie consiste à mesurer, à partir d'électrodes placées sur la peau, l'activité électrique des muscles du visage. Elle permet de détecter une tension musculaire très faible, et donc la présence d'une réaction faciale sans que son expression ne soit forcément visible. En utilisant cette technique, Witvliet et Vrana ont, par exemple montré, une activation des muscles du sourire pendant l'écoute d'une musique joyeuse (Witvliet and Vrana, 1996). En utilisant un EMG, Dimberg a également trouvé que les images de visages heureux suscitaient une activité zygomatique plus importante que les images de visages en colère (Dimberg, 1988).

2.4 Extraction computationnelle de l'information émotionnelle d'une image

La communauté des chercheurs en traitement d'image s'intéresse depuis un certain temps à la mise au point de métriques objectives pour l'évaluation de la qualité d'image (p. ex. (Sheikh et al., 2005, Ke et al., 2006)) et à l'inférence computationnelle de l'information sémantique des images (Datta et al., 2008). Plus récemment, les chercheurs

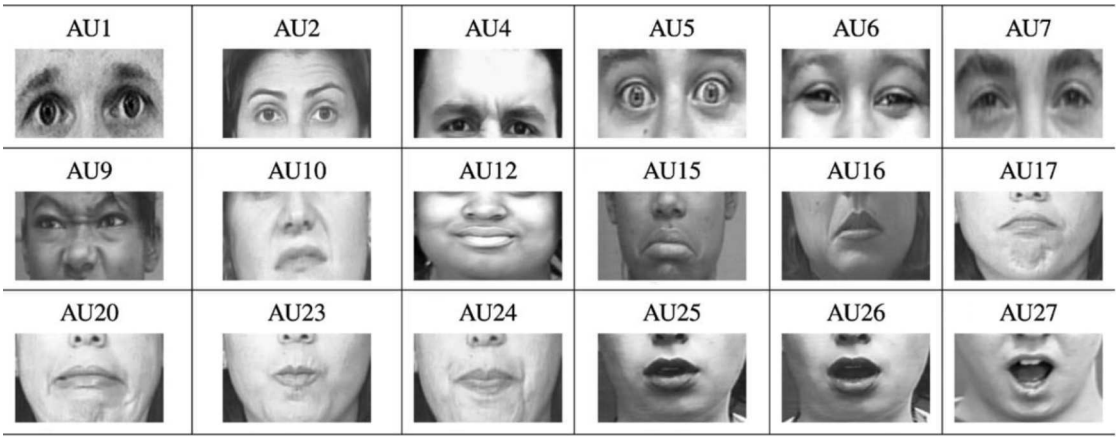


FIGURE 2.6 – Quelques exemple d'unités d'action du *Facial Action Coding System*. Adapté de (Ekman and Friesen, 1978).

se sont appuyés sur ces recherches pour tenter d'associer les émotions véhiculées par les images à leurs caractéristiques intrinsèques (pour une vue d'ensemble, voir (Joshi et al., 2011)). Ces études ont montré qu'il était possible, dans une certaine mesure, d'extraire computationnellement de l'information émotionnelle d'une image. Or, l'information émotionnelle contenue dans une image est étroitement liée à sa mémorabilité (nous en établirons la démonstration dans le chapitre 3). Par conséquent, cette information pourrait potentiellement être mise au service de la prédiction automatique de la mémorabilité d'images. Pour atteindre un tel objectif, il pourra être intéressant de connaître les liens qui unissent l'émotion véhiculée par une image à sa mémorabilité. C'est l'objet du chapitre 7.

Comme pour la mémorabilité, tenter d'extraire l'information émotionnelle d'une image représente un défi ambitieux, puisque nombre d'informations contenues dans l'image (textures, couleurs, sémantique, etc.) sont de potentiels vecteurs d'émotions. En dehors des approches qui ont été développées pour extraire l'information émotionnelle des visages présents dans les images (p. ex. (Valenti et al., 2007, Suja et al., 2016), la plupart des approches visant à extraire l'information émotionnelle des images se sont appuyées sur les caractéristiques visuelles de l'image (couleurs, textures et formes) (Liu et al., 2010b, Lucassen et al., 2010, Wei et al., 2008, Gbèhounou et al., 2012, Machajdik and Hanbury, 2010, Ou et al., 2004). Les couleurs, en particulier, ont fait l'objet d'un certain nombre d'études. Machajdik et col. ont montré, par exemple, un effet direct de la saturation et de la luminosité sur l'arousal, la valence et la dominance en utilisant la base de données d'images IAPS (Machajdik and Hanbury, 2010). L'orientation des différentes lignes contenues dans les images a également été considérée (Liu et al., 2010b). Plus récemment, l'apprentissage profond a également été utilisé dans le cadre de la classification des émotions véhiculées par des images (Chen et al., 2015).

Mettre l'information émotionnelle extraite des images par de tels algorithmes au service de la prédiction de la mémorabilité pourrait conduire, au fur et à mesure que ces algorithmes progresseront, à des progrès indirects de celle-ci.

2.5 Conclusion

Dans ce chapitre, nous avons décidé d'inscrire nos travaux dans une approche dimensionnelle des émotions, et de nous focaliser sur les dimensions d'arousal et de valence. Le relation qui unit ces dimensions présente généralement une forme géométrique en U ou en V, dans les études où des images sont utilisées comme moyen d'induction émotionnelle ; cependant, d'autres formes ont été trouvées. Dans le chapitre 6, nous amènerons de nouveaux résultats et discuterons de ce point plus en détails.

La base IAPS, dont les images ont été évaluées sur les dimensions d'arousal et de valence, nous a semblé être l'outil le plus adapté pour étudier les liens entre la mémorabilité des images et l'émotion qu'elles véhiculent. Nous avons également présenté plusieurs mesures des différentes composantes (cognitive, physiologique et comportementale) de l'émotion, susceptibles d'être utilisées pour l'étude des émotions véhiculées par les images. Certains outils nous ont paru particulièrement intéressants ; en particulier, les échelles SAM, l'oculométrie et l'EEG ont été utilisés dans les travaux présentés dans cette thèse.

Les recherches sur la prédiction computationnelle de la mémorabilité et de l'information émotionnelle des images n'ont, à notre connaissance, jamais été rapprochées. Nous avons émis l'hypothèse qu'il serait profitable d'utiliser l'information émotionnelle extraite des images pour améliorer la prédiction de leur mémorabilité. Nous avons également souligné qu'une connaissance approfondie des liens entre la mémorabilité des images et les émotions qu'elles véhiculent pourrait s'avérer intéressante dans un tel objectif. C'est l'objet du prochain chapitre.

3

L'émotion au cœur des processus mnésiques

Les émotions jouent un rôle important dans la cognition en général, et plus particulièrement dans la mémoire (Dolan, 2002). Les informations suscitant une réaction émotionnelle tendent à être mieux retenues que les informations émotionnellement neutres (Kensinger and Schacter, 2008). Cette interaction entre les processus mnésiques et émotionnels peut être liée à l'évolution humaine : nos mécanismes mnésiques auraient évolué pour nous permettre de nous souvenir des expériences pertinentes pour notre survie ou notre bien-être, qui auront tendance à susciter des émotions (LeDoux, 1998). En effet, selon LeDoux, revivre un évènement antérieur nous permettrait de nous préparer (et éventuellement d'éviter) sa répétition. En particulier, les images émotionnellement chargées tendent à être mieux retenues que les images neutres (p. ex. (Bradley et al., 1992)). L'émotion véhiculée par les images est, par conséquent, un élément qui a tout sa place dans l'étude de leur mémorabilité.

3.1 Effets de l'émotion sur la mémoire lors des différentes étapes du traitement de l'information

L'existence d'une influence de l'émotion sur les processus mnésiques n'est plus en débat (Kensinger and Schacter, 2008). Dans le cadre d'une approche cognitiviste fondée sur le traitement de l'information, on distinguera les influences de l'émotion sur la mémoire lors des différentes étapes du traitement de l'information (l'encodage, le stockage et la

récupération ; voir la figure 3.1). L'émotion influence l'attention visuelle portée à une image (Nummenmaa et al., 2006) et favorise le traitement de l'information émotionnelle (Schwabe and Wolf, 2010), ce qui favorise l'encodage. L'émotion joue également un rôle essentiel dans le processus de consolidation des souvenirs (McGaugh, 2000), qui conduit à la formation de traces mnésiques plus robustes pour l'information émotionnelle. Lors de la récupération de l'information, l'émotion suscitée par un stimulus ou un évènement peut interagir avec l'humeur du participant, et modifier la probabilité qu'il récupère certaines informations (Murray, 1999).

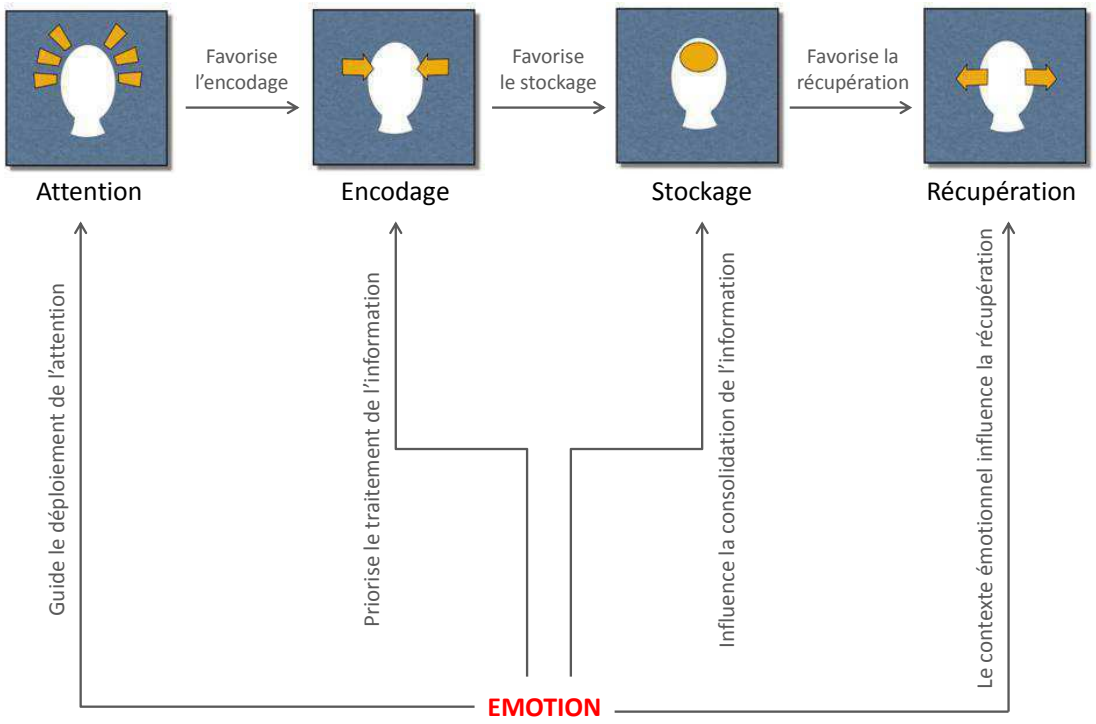


FIGURE 3.1 – Effets de l'émotion sur la mémoire lors des différentes étapes du traitement de l'information.

Les études cherchant à expliquer les effets de l'émotion sur les différentes étapes du traitement de l'information en mémoire ont très souvent adopté une approche dimensionnelle des émotions. En particulier, elles ont principalement porté sur l'arousal et la valence comme facteurs de l'amélioration de la mémoire par l'émotion (Kensinger, 2004). En accord avec l'approche dimensionnelle adoptée dans cette thèse, nous aborderons uniquement les effets de l'arousal et de la valence sur la mémoire lors des différentes étapes du traitement de l'information, sans faire référence aux études ayant adopté une approche catégorielle.

3.1.1 Émotion et encodage mnésique

Au moins deux phénomènes distincts ont été invoqués pour expliquer l'effet de l'émotion sur la mémoire lors de la phase d'encodage : la sélectivité de l'attention et la priorisation des traitements.

L'émotion que véhicule une scène visuelle est un des facteur importants qui vont influencer le déploiement de l'attention visuelle. Ainsi, il a été montré que l'exploration visuelle était sensible à la coloration et la signification émotionnelle de la scène (Christianson et al., 1991, Hermans et al., 1999). De plus, dans une scène visuelle, le fait, pour un élément, de se démarquer par sa coloration émotionnelle permet de sélectionner l'information rapidement et efficacement (p. ex. (Humphrey et al., 2012, Niu et al., 2012, Pourtois et al., 2013)). On conçoit d'ailleurs généralement que les réponses émotionnelles procèdent d'un système adaptatif (Cacioppo and Gardner, 1999). Dans une perspective évolutionniste, la réaction émotionnelle — par exemple, éviter un danger — est cruciale pour la survie ; aussi, l'évolution nous aura rendu sensibles en nous faisant associer une réponse émotionnelle aux informations importantes. Cela explique pourquoi les stimuli suscitant une émotion — en particulier une émotion de peur — sont traités préférentiellement aux stimuli neutres (p. ex. (Blanchette, 2006, Flykt, 2005, Öhman et al., 2001)). Comme pour les scènes visuelles naturelles, les propriétés émotionnelles des images jouent également un rôle important dans l'exploration visuelle et la répartition de l'attention (Quirk and Strauss, 2001). Les propriétés émotionnelles d'une image peuvent aider à organiser son exploration visuelle, en dirigeant l'attention vers les régions les plus informatives (Christianson et al., 1991). De manière plus globale, les images émotionnellement chargées sont généralement plus susceptibles d'attirer l'attention que des images neutres (p. ex. (Nummenmaa et al., 2006). D'autre part, les images suscitant de l'arousal sont plus largement explorées que les images neutres (Quirk and Strauss, 2001). Dans une étude portant sur le traitement préférentiel des scènes émotionnellement chargées, en concurrence pour les ressources attentionnelles avec des scènes émotionnellement neutres, Calvo *et al.* ont montré que les images émotionnellement chargées, mieux reconnues que les images neutres, étaient également associées à une plus haute probabilité d'attirer la première fixation et à des latences de saccades plus courtes (Calvo et al., 2007). En outre, les images suscitant de l'émotion occasionneraient un nombre de fixations plus important que les images neutres (Carniglia et al., 2012). En somme, davantage de ressources attentionnelles sont généralement affectées aux stimuli émotionnellement chargés qu'aux stimuli neutres, contribuant à améliorer la mémoire des premiers relativement aux seconds. Dans la section 11.2 du chapitre 11, nous comparerons le nombre et la durée moyens des fixations occasionnées par des images, et corrèlerons ces mesures au degré de mémorabilité des images et à leurs scores d'arousal et de valence.

Il est également intéressant de mentionner que plusieurs études ont trouvé que la présentation de stimuli émotionnellement chargés, comparativement à la présentation

de stimuli neutres, résultait en une mémoire améliorée pour les détails centraux (i.e. les détails de première importance pour l'apparence ou la signification de l'image) tout en affaiblissant la mémoire des détails périphériques (p. ex. (Burke et al., 1992)). Le *weapon focus effect* (Loftus et al., 1987) est ainsi bien connu en psychologie et en science forensique : les témoins d'un crime ont tendance à se souvenir de manière détaillée de l'arme à feu ou du couteau utilisé, mais peu des autres détails (p. ex. comment était vêtu le criminel). Dans le cadre de l'étude de la mémorabilité des images, il pourrait être intéressant de chercher à prédire la mémorabilité de parties de l'image plutôt que de l'image en entier (comme cela a été fait par (Khosla et al., 2012b)), en s'intéressant à l'émotion véhiculée par ces parties.

Les items émotionnellement chargés sont également plus susceptibles d'être traités que les items neutres lorsque l'attention est limitée, ce qui suggère que le traitement de l'information émotionnelle est prioritaire et/ou plus aisé comparativement au traitement de l'information neutre (Kensinger, 2004). Ce résultat a été montré en utilisant un paradigme dit de *blink attentionnel* (Raymond et al., 1992), suivant lequel deux items cibles sont présentés l'un après l'autre très rapidement dans une séquence de stimuli visuels ; souvent, les participants échoueront à percevoir le second item cible. Plusieurs recherches ont trouvé que, lorsque la seconde cible était un item émotionnellement chargé, la probabilité qu'elle soit perçue était plus élevée (p. ex. (Schwabe and Wolf, 2010, Anderson and Phelps, 2001)), ce qui suggère un traitement facilité pour ce type d'item.

3.1.2 Émotion et rétention mnésique

L'information stockée en mémoire n'est pas immuable. Le stockage dépend de processus actifs : les informations ne sont pas simplement « entreposées », elles sont consolidées et reconstruites pour se combiner aux informations plus anciennes déjà stockées. L'émotion joue une rôle important dans le processus de consolidation de la mémoire, durant la phase de stockage. De nombreuses études ont montré qu'avec l'écoulement du temps, la mémoire des stimuli émotionnellement neutres tendait à diminuer, tandis que la mémoire des stimuli émotionnellement chargés tendait à se maintenir (p. ex. (LaBar and Phelps, 1998)) voire à augmenter (p. ex. (Kleinsmith and Kaplan, 1963)).

Théorie de la consolidation

Selon la théorie de la consolidation, une fois l'information encodée, il faut un certain temps au cerveau pour stabiliser la trace mnésique à travers un processus de consolidation (McGaugh, 2000). Durant la phase de consolidation, les souvenirs sont fragiles et peuvent être facilement altérés ou modifiés : la trace mnésique d'un évènement peut être renforcée ou affaiblie (Brosch et al., 2013).

La plupart des études s'intéressant aux effets de l'émotion sur la phase de rétention mnésique ont porté sur l'effet de l'arousal, considéré comme un facteur essentiel de la consolidation mnésique. L'étude des mécanismes neuraux qui sous-tendent la mémoire de l'information émotionnelle suggère que les hormones de stress liées à l'arousal émotionnel jouent un rôle décisif dans la consolidation de la mémoire (McGaugh, 2000). En cohérence avec cette proposition, plusieurs études ont montré que l'avantage mnésique pour l'information activatrice tendait à être plus important après un délai qu'immédiatement après la phase d'encodage (p. ex. (Eysenck, 1976, Heuer and Reisberg, 1990, LaBar and Phelps, 1998, Revelle and Loftus, 1992)). Quelques études ont même montré que cet avantage mnésique pour l'information activatrice n'apparaissait pas immédiatement après la phase d'encodage (i.e. lorsque la performance de mémoire était mesurée par un test de mémoire immédiat), mais seulement après un délai (i.e. lorsque la performance de mémoire était mesurée par un test de mémoire différé) (Kleinsmith and Kaplan, 1963, Sharot and Phelps, 2004). Dans l'étude de Kleinsmith *et al.* la performance de mémoire des participants était mesurée deux minutes après l'encodage, puis une semaine après, alors que dans l'étude de Sharot et Phelps, la performance de mémoire était mesurée environ trois minutes après l'encodage, puis 24 heures après. L'avantage mnésique pour l'information activatrice causé par l'effet de l'arousal sur la consolidation mnésique apparaitrait donc assez tôt, ce que confirme les études de (Payne et al., 2008a, Sharot and Yonelinas, 2008), dans lesquelles la seconde performance de mémoire des participants était évaluée après un délai de rétention de 12 à 24 heures.

Les techniques d'imagerie cérébrale ont également été mises à profit pour étudier le processus de consolidation mnésique. Par exemple, Steinmetz *et al.* ont étudié la reconnaissance d'images soit activatrices, soit non activatrices, en mesurant l'activation cérébrale durant la phase d'encodage, puis durant deux tâches de reconnaissance, l'une passée 30 minutes après la phase d'encodage, et la seconde 24 heures après (Steinmetz et al., 2012). Les résultats montrent que pour les images activatrices, la correspondance entre l'activation des régions cérébrales durant l'encodage et la performance de mémoire était au moins aussi forte pour le second test de reconnaissance (ayant eu lieu après un délai de 24 heures) que pour le premier test. Au contraire, pour les images neutres, cette correspondance était plus forte lorsque le délai était court. Ces résultats renforcent l'hypothèse d'une consolidation de l'information activatrice, qui s'étendrait au-delà de la première demi-heure en MLT.

Dans la section 1.2.2 du chapitre 1, nous avons suggéré qu'il serait intéressant de mesurer la performance de mémoire de participants quelques minutes après la phase d'encodage mnésique, puis un jour après, pour étudier l'évolution de la mémorabilité des images en MLT. Les études précitées suggèrent que le rôle de l'arousal dans la consolidation peut être mis en évidence par de telles mesures, ce qui permettrait de prendre en compte ce facteur dans l'étude de l'évolution de la mémorabilité des images en MLT. Les intervalles de temps séparant l'encodage des deux tests de mémoire utilisés

dans notre expérience visant à constituer un nouveau jeu de données pour l'étude de la mémorabilité des images, présentée dans le chapitre 5, ont été fixés sur la base de ces constatations.

Nous ajouterons que le sommeil joue un rôle essentiel dans la consolidation. Les traitements mnésiques opérés pendant les phases de sommeil jouent un rôle important dans la formation et l'organisation de nos souvenirs (Stickgold, 2005). Dans l'expérience de (Isola et al., 2011b), qui leur a permis de constituer l'unique base de données actuellement disponible pour l'étude de la mémorabilité d'images, la performance de mémoire à partir de laquelle les scores de mémorabilité ont été calculés était mesurée quelques minutes après l'encodage. Les participants n'ont donc pas dormi entre l'encodage et la récupération. Une littérature de plus en plus conséquente suggère que les effets de l'émotion sur la mémoire sont intensifiés par le sommeil (p. ex. (Payne et al., 2008b, Walker, 2009, Payne and Kensinger, 2011)). Par exemple, Hu *et al.* ont montré que lorsque des images activatrices étaient présentées dans une tâche d'apprentissage avec des images non activatrices, les participants qui dormaient avant la phase de restitution (une tâche de reconnaissance ayant lieu 12 heures plus tard, soit le même jour en condition *Éveil*, soit le lendemain en condition *Sommeil*) étaient meilleurs que les participants qui ne dormaient pas pour reconnaître les images activatrices par rapport aux images non activatrices (Hu et al., 2006). Dans notre étude présentée dans le chapitre 5, nos participant dormaient entre les deux tests de mémoire, ce qui nous a permis d'obtenir une mesure de la performance de mémoire après une phase de sommeil.

3.1.3 Émotion et récupération mnésique

Au-delà des caractéristiques intrinsèques des images, des facteurs individuels sont susceptibles d'influencer la probabilité de récupérer des images en mémoire. En particulier, dans une tâche de reconnaissance d'images, l'humeur[1] des participants au moment de la récupération mnésique pourrait interagir avec l'émotion que véhiculent les images, modifiant la probabilité de reconnaître certaines images. Un effet a notamment été identifié en psychologie, qui suggère une telle possibilité : l'effet de congruence émotionnelle.

La mémoire humaine a tendance à relier des événements avec des significations affectives similaires (Cahill and McGaugh, 1995). L'effet de congruence émotionnelle en est une bonne illustration. Il renvoie à la tendance des individus à récupérer plus facilement des informations en mémoire lorsque leur coloration émotionnelle et celle de leur humeur sont similaires (Lewis and Critchley, 2003). Cet effet a été montré pour des humeurs invoquées et durables (p. ex. les personnes dépressives récupèreraient en mémoire plus de souvenirs de valence négative que positive (Murray, 1999, Watkins et al., 1996) et provoquées (p. ex. l'induction en laboratoire d'une humeur positive par l'écoute d'une

[1] On parlera d'humeur pour indiquer un état affectif durable, que ne modifient pas significativement les émotions, beaucoup plus éphémères.

musique joyeuse augmenterait la probabilité de se remémorer des souvenirs joyeux de son enfance (Martin and Metha, 1997)).

D'autre part, il a parfois été trouvé que la récupération d'une information en mémoire était plus efficace lorsque l'humeur des participants au moment de celle-ci était similaire à leur humeur au moment de l'encodage. Cet effet n'a pas toujours été retrouvé, et son apparition semble beaucoup dépendre de la méthodologie adoptée (Ucros, 1989).

Sans rentrer dans le détail de ces effets, ils peuvent nous porter à nous demander si, dans une tâche de reconnaissance mise en place pour obtenir des scores de mémorabilité pour des images, l'humeur des participants n'influence pas leur performance de mémoire. Pour cette raison, dans notre expérience présentée dans le chapitre 5, nous demandons aux participants d'évaluer leur humeur au cours de la dernière année, et au moment de l'expérience. Dans la section 10.3.3 du chapitre 10, nous proposons un modèle des liens entre plusieurs facteurs individuels — dont l'humeur des participants — et la probabilité de reconnaître une image.

3.2 Effets de l'arousal et de la valence sur la performance de récupération mnésique

Les effet de l'émotion sur les différentes étapes du traitement de l'information mnésique, l'encodage, le stockage et la récupération, vont globalement dans le même sens : la coloration émotionnelle d'une information tend à en renforcer la mémoire. C'est d'ailleurs, comme nous l'avons précédemment évoqué, un fait aujourd'hui bien établi que la mémoire est généralement meilleure pour l'information émotionnelle que pour l'information neutre (Hamann, 2001). La mémoire des évènements émotionnellement chargés est généralement plus vive (p. ex. (Todd et al., 2012)) et plus précise (p. ex. (LaBar and Cabeza, 2006)) que la mémoire des évènements neutres. D'autre part, comme nous l'avons précédemment évoqué, la performance de mémoire est généralement meilleure lorsqu'elle porte sur des stimuli émotionnellement chargés que sur des stimuli neutres ; ce résultat a été observé pour différentes sortes de stimuli, incluant des mots (p. ex. (LaBar and Phelps, 1998)), des histoires (p. ex. (Cahill and McGaugh, 1995)) et des images (p. ex. (Bradley et al., 1992, Ochsner, 2000)). Concernant les images, ce résultat a été observé après des durées de rétention variées : 24 heures (Sharot and Yonelinas, 2008), deux semaines (Comblain et al., 2004), six semaines (Christianson and Fallman, 1990), et un an (Bradley et al., 1992, Dolcos et al., 2005, Weymar et al., 2011).

Suivant l'approche dimensionnelle de l'émotion adoptée dans cette thèse, nous donnons dans la suite de cette section un aperçu des connaissances actuelles sur les effets de l'arousal et de la valence sur la quantité d'information récupérée en mémoire, et sur la qualité de cette récupération. Pour une revue plus détaillée de ces questions, le lecteur

pour se référer à (Kensinger and Schacter, 2008). En préambule de cet aperçu, il est important de rappeler que l'étude de la valence indépendamment de l'arousal est difficile, pour la raison que, très souvent, les items négatifs et positifs sont associés à un arousal plus fort que les items neutres (nous avons évoqué ce point dans la section 2.1.3 du chapitre 2). Dans la plupart des études qui portent sur la comparaison de stimuli positifs, négatifs et neutres, les auteurs s'assurent d'une équipartition de l'arousal dans ces catégories. Cependant, quelques études ont montré que la valence seule pouvait améliorer la mémoire (i.e. des items non activateurs avec une valence positive ou négative étaient mieux rappelés que des items neutres) (p. ex. (Kensinger and Corkin, 2003)).

3.2.1 Effets de l'arousal et de la valence sur la quantité d'informations récupérées

Dans l'acception qui lui est donnée ici, la quantité d'information récupérée renvoie au nombre d'items rappelés ou reconnus dans un test de mémoire, après avoir été vus durant une phase d'apprentissage préalable. Les taux de récupération sont généralement plus élevés pour les stimuli activateurs que pour les stimuli non activateurs (p. ex. (Christianson, 1992, Mather and Sutherland, 2011). Ce résultat a été répliqué en utilisant des paradigmes expérimentaux et des matériels variés (Kensinger and Schacter, 2008, Madan et al., 2012). Par exemple, il a été montré par Bradley *et al.* pour des images issues de l'IAPS (Bradley et al., 1992). De la même manière, il est bien établi que les taux de récupération sont généralement plus élevés pour les stimuli positifs et négatifs que pour les stimuli neutres (voir la revue de (Buchanan and Adolphs, 2002, Christianson, 1992)). Toutefois, un point peu clair concerne la probabilité relative des stimuli négatifs et positifs d'être récupérés en mémoire (Kensinger and Schacter, 2008). Par exemple, dans (Bradley et al., 1992), les taux de récupération des images positives et négatives étaient équivalents, alors que dans (Charles et al., 2003), les taux de récupération étaient plus élevés pour les images négatives que positives (l'avantage relatif des images négatives tendait cependant à baisser avec l'âge des participants, regroupés par catégories : adultes jeunes, d'âge moyen ou âgés). Selon (Kensinger and Schacter, 2008), dans les études portant sur des images, ce dernier pattern (taux de récupération plus élevé pour les stimuli négatifs que positifs) serait le plus courant. Nous apporterons des résultats pour nourrir le débat autour de cette question, dans la section 7.2 du chapitre 7.

3.2.2 Effets de l'arousal et de la valence sur la qualité des informations récupérées

Tous les souvenirs ne sont pas créés égaux : lors de certaines récupérations mnésiques, nous nous sentons comme transportés dans le temps en refaisant l'expérience d'un évènement antérieur, et notre mémoire semble contenir une grand nombre de détails sur

l'endroit et le moment où l'évènement est advenu ; d'autres fois, nous reconnaissons quelque chose que nous avons vu auparavant, mais notre mémoire ne contient pas d'information sur le contexte de cette rencontre (Kensinger and Schacter, 2008). Nous entendons par qualité d'une information récupérée, la puissance de la reviviscence du souvenir, qui s'appuie sur les éléments contextuels associés à l'évènement mis en mémoire. La grande majorité des études qui ont examiné les effets de l'émotion sur la qualité de la récupération mnésique se sont concentrées sur la dimension d'arousal (Kensinger and Schacter, 2008). Selon ces auteurs, il est bien établi que la qualité de la récupération mnésique est généralement meilleure pour les stimuli activateurs que pour les stimuli non activateurs. En particulier, ce résultat a été trouvé pour des images (Ochsner, 2000, Sharot et al., 2004). Quelques études ont également porté sur l'effet de la valence sur qualité de la récupération mnésique. Étant donnée la difficulté à trouver un matériel positif ou négatif non activateur, les chercheurs ont généralement opposé la qualité de la récupération des stimuli négatifs activateurs à celle des stimuli positifs activateurs. Les résultats de ces études suggèrent que la qualité de la récupération tend à être meilleure pour les stimuli négatifs que positifs (Kensinger and Schacter, 2008).

Les travaux présentés dans cette thèse portent uniquement sur les effets de l'arousal et de la valence sur la quantité d'information récupérée, mesurable par une tâche de reconnaissance du type de celle employée par (Isola et al., 2011b). Toutefois, il nous paraît intéressant de noter qu'il est possible de s'intéresser à la qualité de l'information récupérée dans le cadre d'une tâche de reconnaissance, en utilisant des paradigmes tels que les paradigmes *what-where-when* et *Remember-Know*, présentés dans la section 1.2.3 du chapitre 1. Par exemple, Ochsner a étudié les états de conscience accompagnant la reconnaissance d'images de l'IAPS en utilisant le paradigme *Remember-Know* (Ochsner, 2000). Il a trouvé que les images négatives occasionnaient plus de réponses *Remember* que de réponses *Know*, et a trouvé le pattern inverse pour les images positives, ce qui suggère que la qualité de la récupération mnésique tend à être meilleure pour les images négatives que positives. On peut imaginer élargir un jour l'étude de la mémorabilité des images en informatique à la qualité de la récupération mnésique.

3.3 Conclusion

Dans ce chapitre, nous avons présenté des connaissances produites en psychologie qui établissent l'existence d'un lien entre l'émotion et la mémoire humaine. En particulier, de nombreuses études indiquent que l'arousal et la valence véhiculés par les images sont des facteurs importants à prendre en compte dans l'étude de leur mémorabilité. Pour étudier conjointement ces facteurs, il est important de distinguer l'étape du traitement de l'information mnésique lors de laquelle l'effet de l'émotion sur la mémoire se produit. Les études portant sur l'encodage renforcent notre intérêt pour l'attention visuelle, initialement suscité par une étude de Mancas et Le Meur qui suggère que l'extraction de

caractéristiques des images liées à l'attention visuelle est pertinente dans le cadre de la prédiction informatique de leur mémorabilité (Mancas and Le Meur, 2013). Les études portant sur le stockage suggèrent que la mémorabilité des images activatrices pourrait évoluer en MLT, différemment des images non activatrices, en raison du rôle de l'arousal dans le processus de consolidation. Par conséquent, ce facteur parait important pour investiguer la question d'un éventuel changement dans le temps du degré — et surtout de l'ordre — de mémorabilité des images, qui jetterait une nouvelle lumière sur les scores de mémorabilité obtenus par (Isola et al., 2011b). Pendant la récupération, l'humeur dans laquelle se trouvent les participants pourrait favoriser la récupération mnésique des images émotionnellement chargées congruentes — et donc augmenter globalement la probabilité de récupération des images émotionnellement chargées, sans changer la probabilité de récupération des images neutres. Cette possibilité rappelle l'importance des facteurs individuels dans l'étude de la mémoire.

A notre connaissance, ces différents points n'ont jamais été abordés dans l'étude de la mémorabilité des images en informatique, comme nous le verrons dans le prochain chapitre, qui porte spécifiquement sur ce domaine d'étude.

4

Vers une prédiction computationnelle de la mémorabilité des images

L'étude de la mémorabilité des images en informatique a commencé assez récemment ; à notre connaissance, la première étude consacrée à ce sujet date de 2011 (Isola et al., 2011b). Cette étude a reposé sur le constat initial que certaines images sont intrinsèquement plus mémorables que d'autres, indépendamment du contexte ou de l'observateur (Isola et al., 2011a). Son objectif était de découvrir des caractéristiques spécifiques aux images mémorables, pour développer un modèle représentatif de ce type d'image. Cette première tentative, qui s'est appuyée sur des algorithmes d'apprentissage pour inférer le degré de mémorabilité d'images de caractéristiques de bas niveau de celles-ci, a montré qu'un tel objectif était atteignable (Isola et al., 2011b). Depuis, plusieurs études ont porté sur la question (p. ex. (Khosla et al., 2012b, Mancas and Le Meur, 2013, Kim et al., 2013, Celikkale et al., 2013, Isola et al., 2014, Bylinskii et al., 2015a, Celikkale et al., 2015, Wang et al., 2015, Lahrache et al., 2016).

L'étude de la mémorabilité des images avec un objectif de prédiction est interdisciplinaire : la définition de la mémorabilité, sa mesure et la compréhension des facteurs susceptibles de l'influencer relèvent de la psychologie, et le développement de modèles prédictifs relève de l'informatique. Cette étude nécessite, par conséquent, une approche transversale. Dans ce chapitre, nous donnons un aperçu des travaux portant sur la mémorabilité des images en informatique, à la lumière des connaissance produites par la psychologie, présentées dans les chapitres précédents.

4.1 Les facteurs qui influencent la mémorabilité d'une image

Les raisons qui font qu'une image est mémorable sont multiples. D'une part, les caractéristiques intrinsèques des images jouent un rôle dans leur mémorabilité (Isola et al., 2014, Lahrache et al., 2016). Certaines images peuvent contenir des éléments qui sont familiers à la personne qui les regarde ; par exemple, elles peuvent contenir des amis, ou des monuments qu'elle a visité. D'autres images qui ne contiennent aucun monument ou personne reconnaissables peuvent également être hautement mémorables (Brady et al., 2008). C'est de ce dernier type d'images — qui constitue la base de données de (Isola et al., 2011b), la seule disponible à ce jour, à notre connaissance, pour l'étude de la mémorabilité des images — que les études existantes se sont attachées à prédire la mémorabilité.

Des facteurs extrinsèques de l'image sont également susceptibles d'influencer sa mémorabilité. Le contexte de présentation d'une image peut avoir un impact sur sa mémorabilité (Bylinskii et al., 2015b). La mémorabilité peut également varier en fonction de l'observateur et de l'état dans lequel il se trouve (par exemple, selon son humeur, comme nous l'avons expliqué dans la section 3.1.3 du chapitre précédent). Les facteurs extrinsèques ont cependant largement été laissés de côté jusqu'à aujourd'hui, la plupart des études existantes s'étant concentrées sur la prédiction d'un score de mémorabilité moyen pour les images, sans chercher à l'adapter au contexte ou à l'observateur.

4.1.1 Les caractéristiques des images liées à leur mémorabilité

En traitement d'images, on distingue généralement les caractéristiques de bas et de haut niveau des images. Le terme de « bas niveau » est généralement utilisé pour désigner les caractéristiques de l'image qui sont directement liées aux valeurs des pixels des images (Szummer and Picard, 1998) ; par exemple, la texture, la forme, la couleur, la localisation spatiale, etc. (Upadhyaya and Dixit, 2016). Ces caractéristiques de bas niveau n'ont pas de signification telle qu'un humain peut en donner en interprétant l'image (Xue et al., 2013). Le terme « haut niveau » renvoie à l'interprétation (au sens large) des images par un humain (Upadhyaya and Dixit, 2016). Des caractéristiques de haut niveau sont, par exemple, le type de scène représentée par l'image ou l'émotion qu'elle véhicule. D'autre part, on distingue également les caractéristiques globales, qui sont calculées sur l'image dans sa globalité, des caractéristiques locales, qui sont calculées sur des régions ou objets de l'image (Vassilieva, 2009).

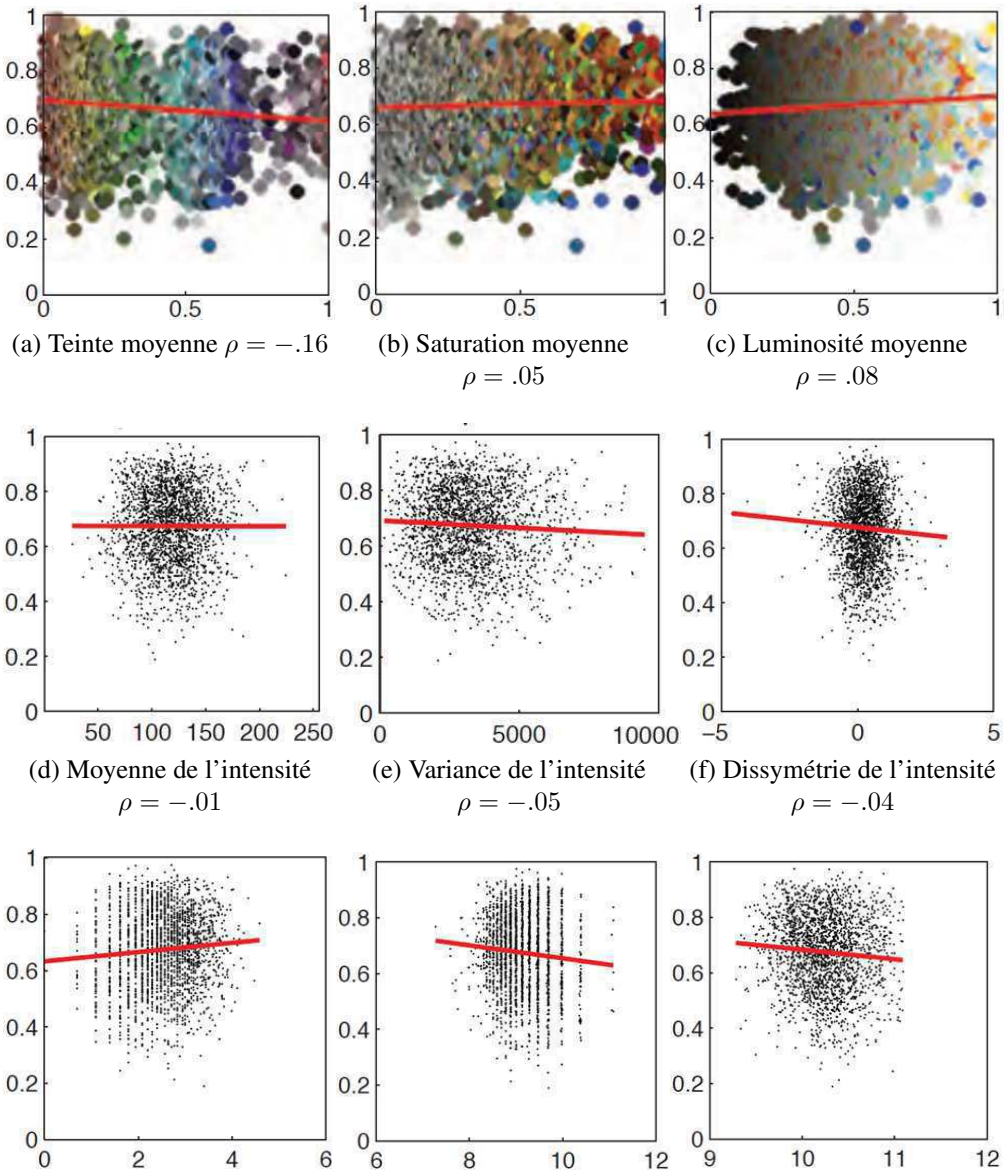
Caractéristiques de bas niveau des images

Une des première question à avoir été posée est la suivante : est-ce que des caractéristiques de bas niveau des images, prises isolément, suffisent pour déterminer si une image est mémorable ? (Isola et al., 2011b) Pour répondre à cette question, Isola *et al.* ont calculé les coefficients de corrélation entre les scores de mémorabilité d'images (calculés à partir des résultats obtenus à une tâche de reconnaissance) et plusieurs caractéristiques de bas niveau de ces mêmes images (voir les graphiques (a), (b), (c), (d), (e) et (f) de la figure 4.1). Leurs résultats montrent que la mémorabilité d'une image a tendance à baisser lorsque la moyenne des valeurs de teinte de l'image (en degrés ; transition du rouge au vert, au bleu, au violet) augmentent (Isola et al., 2011b, Isola et al., 2014). En revanche, la saturation et la luminosité moyennes de l'image, ainsi que les trois premiers moments (i.e. moyenne, variance et coefficient de dissymétrie) de l'histogramme d'intensité des pixels (voir la figure 4.2) étaient très faiblement corrélés à la mémorabilité. Les auteurs rapprochent ce résultat de ceux obtenus dans une étude précédente (Konkle et al., 2010), qui suggère que les caractéristiques perceptuelles ne sont pas retenues en mémoire à long terme. Le fait que la teinte soit corrélée à la mémorabilité pouvant être expliqué, selon Isola et al., par le fait que les scènes d'extérieur — souvent bleues et vertes — ont globalement une mémorabilité plus faible que les scènes d'intérieur ou représentant des visages humains, généralement de teintes plus chaudes (Isola et al., 2011b).

Plus généralement, les modèles de prédiction combinent de nombreuses caractéristiques (de bas et de haut niveaux) des images, et prennent en compte leurs interactions (p. ex. (Khosla et al., 2012b, Mancas and Le Meur, 2013, Kim et al., 2013, Isola et al., 2014, Lahrache et al., 2016)). La complexité des modèles utilisés (voir section 4.3.1), à *boîte noire*, rend cependant difficile l'interprétation de la relation entre chacune de ces caractéristiques et la mémorabilité des images. Nombre de ces caractéristiques ont été précédemment utilisées pour résoudre d'autres types de problème de vision par ordinateur ; par exemple, (Isola et al., 2011b) ont utilisé, pour leur modèle de prédiction de la mémorabilité des images, la liste de caractéristiques proposée par (Xiao et al., 2010) pour la reconnaissance de scènes. Par conséquent, les auteurs travaillant sur la mémorabilité des images en vision par ordinateur peuvent bénéficier d'algorithmes déjà bien connus dans leur domaine de recherche ; par exemple, les algorithmes SIFT (Lowe, 2004), HOG2x2 (Dalal and Triggs, 2005), SSIM (Shechtman and Irani, 2007), GIST (Oliva and Torralba, 2001), etc., qui permettent d'extraire des images des descripteurs variés.

Statistiques portant sur les objets

Les objets présents dans les images ont une influence importante sur leur mémorisation (Koutstaal and Schacter, 1997, Konkle et al., 2010). Les caractéristiques des images



(a) Teinte moyenne $\rho = -.16$ (b) Saturation moyenne $\rho = .05$ (c) Luminosité moyenne $\rho = .08$

(d) Moyenne de l'intensité $\rho = -.01$ (e) Variance de l'intensité $\rho = -.05$ (f) Dissymétrie de l'intensité $\rho = -.04$

(g) Log du nombre d'objets dans l'image $\rho = .07$ (h) Log de la taille moyenne des objets dans l'image $\rho = -.06$ (i) Log de la taille maximum d'un objet dans l'image $\rho = -.09$

FIGURE 4.1 – Quelques caractéristiques des images et de statistiques calculées sur les objets des images, et leurs corrélations (ρ de Spearman) avec les scores de mémorabilité des images (en ordonnées). Comme souligné dans (Isola et al., 2014), d'où est tirée la figure adaptée ici, les corrélations sont relativement faibles : ces seules caractéristiques, prises isolément, ne permettent pas de faire de bonnes prédictions de la mémorabilité.

FIGURE 4.2 – Histogramme d'intensité de l'image 1595 de l'IAPS. Cet histogramme de la distribution des intensités de l'image associe à chaque valeur d'intensité le nombre de pixels dans l'image prenant cette valeur. (Pour une image en niveau de gris, les pixels prennent généralement une valeur entre 0 et 255.)

liées aux objets ont, pour cette raison, dès le départ suscité l'intérêt des chercheurs s'intéressant à la mémorabilité des images en informatique (Isola et al., 2011b). Pour déterminer (entre autres) si certaines statistiques de base portant sur les objets ont, lorsqu'elles sont considérées isolément, une influence significative sur la mémorabilité des images, Isola *et al.* ont utilisé un outil mis à disposition des chercheurs, *LabelMe* (Russell et al., 2007). Cet outil, représenté dans la figure 4.3, permet de découper les images en différentes régions correspondant à un objet particulier, et d'attribuer à ces régions une étiquette descriptive (p. ex. « personne », « fenêtre », « arbre », etc.). Isola *et al.* ont ensuite calculé les coefficients de corrélation entre le nombre d'objets dans l'image, la taille moyenne des objets dans l'image et la taille maximum d'un objet dans l'image, et les scores de mémorabilité des images (voir les graphiques (g), (h) et (i) de la figure 4.1). Les corrélations, très faibles, suggèrent qu'aucune de ces statistiques portant sur les objets ne permet à elle seule de faire de bonnes prédictions de la mémorabilité.

Attributs sémantiques des scènes et objets

Selon (Isola et al., 2011b), les résultats de leur étude suggèrent que la simple considération des objets à travers des statistiques, sans prendre en compte leur dimension sémantique, n'est pas efficace pour prédire la mémorabilité des images. Le sens d'une scène et des objets représentés dans les images jouent un rôle important dans leur mémorisation (Konkle et al., 2010, Koutstaal and Schacter, 1997). Il est donc important de s'intéresser aux caractéristiques de haut niveau des images dans un cadre de prédiction de leur mémorabilité. Isola *et al.* ont, en particulier, étudié 127 attributs sémantiques[1] de photographies, portant sur l'agencement spatial de la scène (p. ex. ouverte,

[1]Le détail des caractéristiques utilisées peut être trouvé dans (Berg et al., 2012), où les auteurs proposent un certain nombre de caractéristiques, dont ils étudient l'influence sur l'importance du contenu

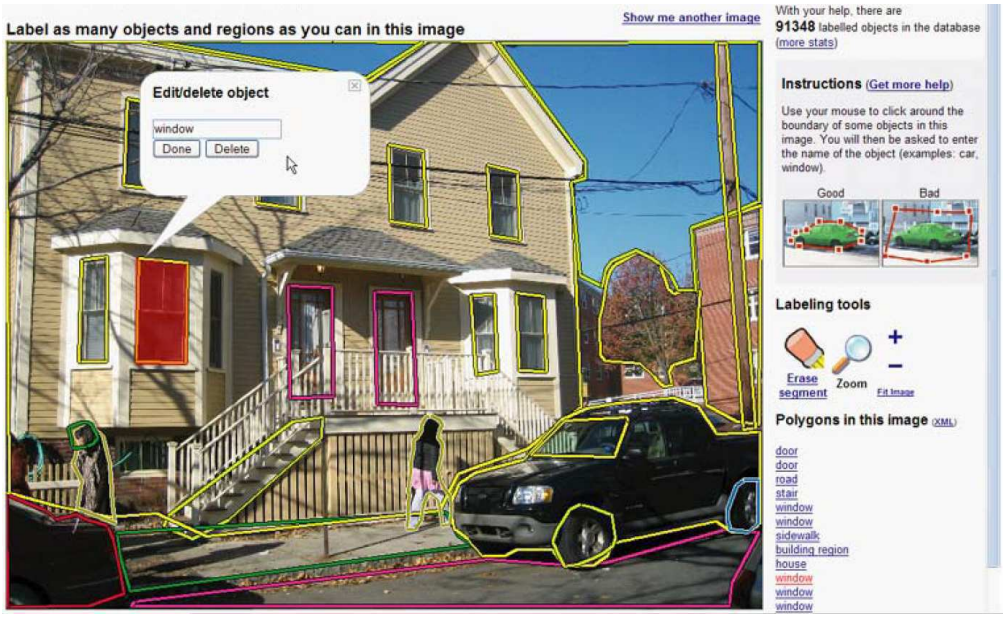


FIGURE 4.3 – Capture d'écran de LabelMe (Russell et al., 2007), utilisé par (Isola et al., 2011b) pour le découpage et l'annotation des objets contenus dans les images de leur base de données.

encombrée), l'esthétique, la dynamique de la scène (p. ex. statique, dynamique, objets en mouvements), la localisation (p. ex. place célèbre), l'émotion associée à l'image (p. ex. plaisante, drôle), l'action (p. ex. des personnes en conversation, qui s'assoient) et l'apparence des personnes (p. ex. genre, habits) (Isola et al., 2014). Leurs résultats montrent que l'apprentissage automatique de liens entre les attributs sémantiques des photographies et leur mémorabilité est un moyen efficace de prédire cette dernière.

Dans un cadre de prédiction automatique, la difficulté principale réside dans le fait que l'extraction par algorithmes des attributs sémantiques des images est complexe. Cependant, les progrès sont rapides, en particulier grâce au développement des techniques d'apprentissage profond, et à la disponibilité de larges bases de données d'images annotées, rendue possible grâce à l'internet. La progression dans la prédiction de la mémorabilité sera probablement étroitement liée à celle de notre capacité à extraire efficacement les caractéristiques de haut niveau des images.

---

des images, telle qu'elle est perçue et jugée par des être humains.

Caractéristiques de l'image liées à l'attention visuelle

Récemment, Mancas et Le Meur (Mancas and Le Meur, 2013) se sont intéressés aux caractéristiques de bas niveau des images liées à l'attention visuelle[2] pour en prédire la mémorabilité. Plus précisément, ils ont isolé deux caractéristique des images utiles dans un tel objectif : la *couverture de saillance* des images et le contraste des structures présentes dans l'image. Ces deux caractéristiques sont extraites des images à l'aide de modèles computationnels d'attention visuelle : elles sont, plus précisément, calculées à partir des cartes de saillance[3] générées par ces modèles.

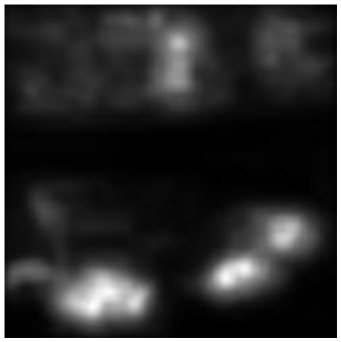
Pour étudier la couverture de saillance des images, Mancas et Le Meur ont sélectionné dans la base de données de (Isola et al., 2011b) des images associées à des scores de mémorabilité, qu'ils ont réparties dans trois classes : $C1$ pour les images hautement mémorables, $C2$ pour les images moyennement mémorables, et $C3$ pour les images peu mémorables. Ils ont ensuite utilisé des modèles d'attention visuelle pour générer des cartes de saillance pour chacune de ces images. Un exemple de carte de saillance générée par le modèle RARE (Riche et al., 2013) est donné dans la figure 4.4. (Nous reviendrons plus en détails sur les modèles d'attention visuelle et les cartes de saillance dans le chapitre 11.) À partir des cartes de saillance, ils ont ensuite calculé la couverture moyenne de saillance des images de chaque classe, qui décrit la distribution spatiale de la densité de saillance des images, en moyennant les cartes de saillance de l'ensemble des images de la classe. Les résultats (obtenus à partir des cartes de saillance générées par le modèle RARE) sont illustrés par la figure 4.5. Ils montrent une différence significative de couverture de saillance moyenne entre les images de la classe $C1$ (i.e. les images les plus mémorables) et les images des classes $C2$ et $C3$. Dans la classe $C1$, la couverture de saillance des images tend à être plus faible que pour les deux autres classes, ce qui, selon les auteurs, suggère que les images de cette classe tendent à présenter une unique région d'intérêt, par rapport aux images (moins mémorables) des deux autres classes, qui tendent soit à ne pas présenter de régions d'intérêt précises, soit à en avoir plusieurs.

Pour étudier le contraste des structures de l'image, Mancas et Le Meur ont appliqué des filtre passe-bas aux images, l'application successive des ces filtres éliminant progressivement les détails de l'image (voir la figure 4.6). Pour évaluer l'influence du filtrage passe-bas, ils ont calculé les corrélations entre l'image initiale et les images filtrées (une corrélation correspond à la moyenne des corrélations des composants RGB ou *Red, Green, Blue*). L'idée est que les images présentant des structures fortement contras-

[2]Ces caractéristiques sont liées à l'attention visuelle dans le sens où elles ont une influence sur le déploiement de l'attention visuelle des observateur des images, qui est liée à la mémorabilité des images, comme l'ont montré (Mancas and Le Meur, 2013).

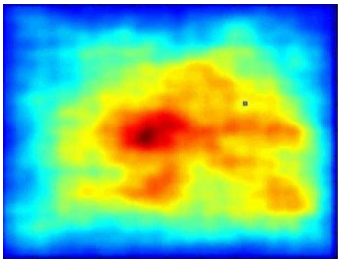
[3]En vision par ordinateur, les modèles d'attention visuelle sont principalement basés sur le concept de cartes de saillance (Riche et al., 2013). Un modèle génère une carte de saillance à partir d'une image, qui prédit les endroits de cette image où un observateur humain est le plus susceptible de porter son attention.

(a) Image originale

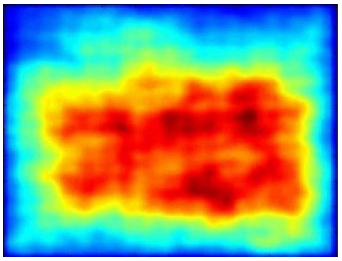


(b) Carte de saillance

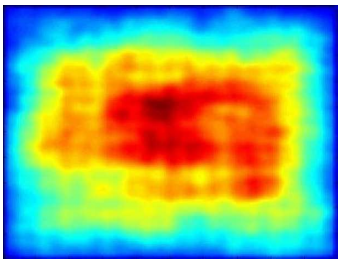
FIGURE 4.4 – Exemple de carte de saillance générée par un modèle d'attention visuelle, le modèle RARE (Riche et al., 2013). Tirée de (Mancas and Le Meur, 2013).



(a) $C1$



(b) $C2$



(c) $C3$

FIGURE 4.5 – Couvertures moyennes des images des classes $C1$, $C2$ et $C3$ par des zones saillantes, calculées à partir des cartes de saillance générées par le modèle RARE (Riche et al., 2013). (Tiré de (Mancas and Le Meur, 2013).)

tées tendent à être plus résistantes aux filtrages passe-bas (ce qui se traduit par une plus forte corrélation) que les images présentant des détails fins et des fonds chargés. Leurs résultats montrent des différences significatives de contraste moyen des structures des images entre les images les moins mémorables et les images les plus mémorables.

FIGURE 4.6 – Filtres passe-bas appliqués successivement à une image. De gauche à droite, l'image originale, puis les images filtrées. (Tiré de (Mancas and Le Meur, 2013).)

En combinant ces caractéristiques de bas niveau des images liées à l'attention visuelle avec des caractéristiques proposées par (Isola et al., 2011b), Mancas et Le Meur ont trouvé que leurs caractéristiques pouvaient avantageusement remplacer plusieurs des caractéristiques précédentes, leur permettant d'atteindre une performance de prédiction légèrement meilleure avec un nombre de caractéristiques significativement plus bas.

4.1.2 Les facteurs extrinsèques

En plus des caractéristiques intrinsèques des images, d'autres facteurs, extrinsèques, sont susceptibles d'influencer la mémorabilité d'une image ; en particulier, les facteurs liés au contexte de présentation des images, et les facteurs individuels.

Nous avons tendance à nous souvenir des choses qui se distinguent du contexte local de leur présentation (p. ex. (Rawson and Van Overschelde, 2008, Schmidt, 1985)). Pour cette raison, Isola *et al.* ont mesuré la corrélation entre la fréquence du contenu des images de leur base de données et leurs scores de mémorabilité (Isola et al., 2011b). Plus précisément, ils se sont intéressés à la fréquence des images contenant un objet particulier, à la fréquence des objets eux-mêmes, et à la fréquence du type de scène représentée dans l'image[4] (p. ex. « forêt », « chambre »). L'analyse de leurs résultats montre une corrélation négative significative ($\rho = -0.13$) entre la fréquence du type de scène et les scores de mémorabilité des images. Ce résultat suggère qu'une scène moins représentée dans le contexte dans lequel les images sont visionnées tend à être plus mémorable. En revanche, les corrélations entre, d'un côté, la fréquence des images contenant un objet particulier, et la fréquence des objets eux-mêmes, dans la base de données, et, de

[4]La catégorisation des scènes représentées par les images a été réalisée par les auteurs de la base SUN (Xiao et al., 2014), d'où proviennent les images utilisées par Isola *et al.*.

l'autre, la mémorabilité des images, étaient très faibles ($\rho = -0.05$ et $\rho = 0.01$, respectivement). Dans leur étude, l'objectif des auteurs en réalisant une telle analyse était de s'assurer qu'un éventuel biais de fréquence dans leur base de données n'expliquait pas les résultats de mémorabilité obtenus pour leurs images. Les auteurs recommandent, pour tester de plus subtiles interactions entre la mémorabilité et le contexte de présentation des images, de mesurer la mémorabilité sur de nouveaux jeux d'images, ce qui permettrait de mesurer à quel point leurs résultats sont généralisables.

Plus récemment, Bylinskii *et al.* ont proposé une méthode pour prendre en compte le contexte de manière automatique dans la prédiction de la mémorabilité des images (Bylinskii et al., 2015b). Cette méthode permet d'estimer dans quelle mesure chacune des images présentées une tâche de reconnaissance est distincte des autres, sur la base d'un ensemble de caractéristiques intrinsèques des images algorithmiquement extractibles. Leurs résultats montrent que plus une image est distincte de son contexte de présentation, plus elle tend à être mémorable. La distinctivité en matière d'émotion véhiculée par les images n'a cependant, à notre connaissance, jamais été prise en compte dans l'étude de la mémorabilité des images en informatique ; c'est l'objet principal du chapitre 9.

D'autre part, il existe une part de subjectivité dans la mémorabilité (Hunt and Worthen, 2006) : une image mémorable pour une personne ne l'est pas nécessairement pour une autre. Nous le confirmons dans la section 7.1 du chapitre 7, où nous posons la question de la variabilité inter-sujet de la mémorabilité des images. Les facteurs individuels n'ont, à notre connaissance, jamais été étudiés dans un cadre de prédiction de la mémorabilité des images, probablement à cause de la complexité qu'ajoute cette dimension aux modèles et à l'absence de base de données qui donnent accès aux détails des résultats. Dans le chapitre 10, nous proposons un modèle de l'influence de plusieurs facteurs individuels (âge, genre, etc.) sur la probabilité de reconnaître une image. D'autre part, nous avons prévu de rendre disponible à la communauté le détail des données obtenues dans notre expérience visant à constituer une nouvelle base de données pour la mémorabilité d'image, présentée dans le chapitre 5.

4.2 La mémorabilité des images dans les travaux existants

L'étape visant à collecter des scores de mémorabilité pour les images est d'une grande importance. La mémoire est un objet immatériel, multiple, et susceptible, comme nous l'avons établi, de subir l'influence de nombreux facteurs. Le type de test utilisé pour mesurer la mémoire et les modalités de sa mise en œuvre déterminent ce que seront les scores de mémorabilité. En particulier, la durée de présentation des images détermine, comme nous l'avons souligné, la surcharge potentielle de la mémoire à court terme, susceptible d'altérer la mémorisation, et est susceptible d'influencer le déploiement de

l'attention visuelle. D'autre part, la durée de rétention mnésique qui sépare l'encodage des images de leur récupération détermine le type de mémoire — mémoire à court ou à long terme — sur laquelle porte le test ; et les processus de consolidation sont susceptibles de modifier les souvenirs des images durant cet intervalle de temps. Les modèles de prédiction de la mémorabilité des images existants ont en commun un apprentissage automatique réalisé à partir d'images associées à des scores de mémorabilité issus d'une même base de données, proposée par (Isola et al., 2011b). La manière dont les scores de mémorabilité ont été obtenus mérite, par conséquent, un intérêt particulier, puisqu'elle détermine ce qu'est la mémorabilité que l'on cherche à prédire en informatique — et ce qu'elle n'est pas.

4.2.1 La méthode employée pour mesurer la mémorabilité des images

La première étude ayant, à notre connaissance, porté sur l'étude de la mémorabilité des images en informatique (Isola et al., 2011b), a eu une influence considérable sur les études subséquentes qui se sont inscrites dans la continuité de ce travail. En particulier, ces études ont utilisé les images issues de la base de données proposée par ces auteurs avec leurs scores de mémorabilité (Isola et al., 2011a, Khosla et al., 2012a, Khosla et al., 2012b, Mancas and Le Meur, 2013, Isola et al., 2014, Celikkale et al., 2013, Celikkale et al., 2015, Kim et al., 2013), ou les auteurs ont fait noter la mémorabilité d'images en employant une méthode similaire à celle de ces auteurs (Khosla et al., 2013, Bylinskii et al., 2015b). Cette unique *vérité terrain*, ou encore le concept de mémorabilité qui en découle, n'ont pas, à notre connaissance, été véritablement réinterrogés.

Les scores de mémorabilité pour les images ont été obtenus par (Isola et al., 2011b) par le moyen d'un « jeu de mémoire » — en fait, une tâche de reconnaissance dans laquelle l'encodage et la récupération des différentes images sont entrelacés (voir la figure 4.7). Les participants avaient accès au jeu via une plateforme de crowdsourcing (voir la section 4.2.2). Durant le jeu, ils voyaient une séquence d'images, affichées chacune durant une seconde, et séparées entre elles par un espace de $1,4$ seconde. Leur tâche était de presser la barre d'espacement dès qu'ils reconnaissaient une image qu'ils avaient précédemment vue dans le jeu. Seule une partie des images — les images cibles — étaient présentées deux fois (en laissant de côté les images utilisées pour le test de vigilance ; voir la section 4.2.2) : une fois pour l'encodage et une fois pour la récupération. Les autres images étaient des images de remplissage, qui corsaient le test et permettaient d'éloigner les premières occurrences des images cibles de leurs répétitions. Cet éloignement était de 91 à 109 images, soit de une minute et 31 secondes à une minute et 49 secondes ; c'est donc une performance de mémoire à long terme que la tâche de reconnaissance mesurait. Lorsqu'un participant appuyait sur la barre d'espacement, un retour — matérialisé par une croix verte ou rouge — lui signifiait soit qu'il avait reconnu une image effectivement répétée, soit qu'il avait commis une fausse alarme, c'est-à-dire qu'il avait reconnu une image vue pour la première fois. Les images étaient réparties

en niveaux de 120 images chacun, d'une durée de 4 minutes et 48 secondes. A la fin de chaque niveau, le taux de réponses correctes et incorrectes du participant pour le niveau était affiché. Un participant pouvait compléter au maximum 30 niveaux (et pouvait quitter le jeu à n'importe quel moment).

A la suite de l'expérience, Isola *et al.* ont calculé un score de mémorabilité pour chacune des 2222 images cibles (les 8220 autres images étant utilisées uniquement pour le remplissage), toutes sélectionnées aléatoirement dans la base d'images SUN (Xiao et al., 2010). Un score de mémorabilité pour une image correspond au taux moyen de reconnaissance de cette image, calculé à partir des réponses de 78 participants en moyenne. La figure 4.8 est constituée de quelques unes des images de la base de données de (Isola et al., 2011b), triées par ordre décroissant de leur mémorabilité. Le score moyen de mémorabilité est de $67,5\%$ ($\sigma = 13,6\%$) pour l'ensemble des 2222 images. Le taux de fausses alarmes, calculé sur les images de remplissage, était de $10,7\%$ ($\sigma = 7,6\%$).

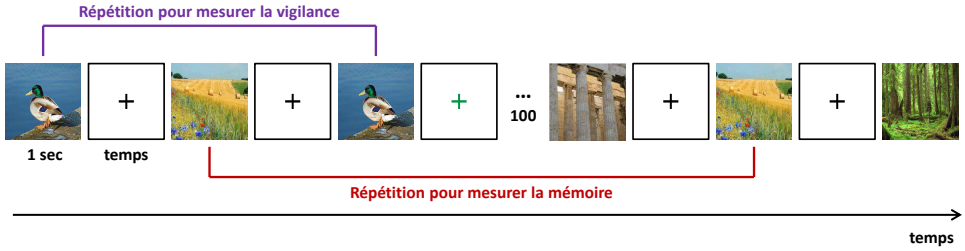


FIGURE 4.7 – La tâche de mémoire proposée en crowdsourcing pour collecter des scores de mémorabilité pour des images numériques (adaptée de (Isola et al., 2011b)).

4.2.2 Des scores obtenus en crowdsourcing

Comme nous l'avons évoqué, l'expérience de (Isola et al., 2011b), à partir des résultats de laquelle les scores de mémorabilité des images de leur base de données ont été obtenus, a été proposée en crowdsourcing. Cette technique qui, dans un cadre de recherche scientifique, consiste généralement à verser une expérience sur une plateforme en ligne et à recruter des utilisateurs d'Internet comme participants, permet d'obtenir rapidement un grand nombre de participants. Elle est généralement plus économique que les études menées en laboratoire. Cependant, le contrôle des facteurs susceptibles d'influencer la tâche est plus difficile, d'où une utilisation généralisée de tests de fiabilité pour s'assurer que les participants réalisent bien la tâche demandée. Le test de vigilance (voir figure 4.7) mis en place par (Isola et al., 2011b) avait cette destination : certaines des images de remplissage étaient répétées quelques secondes après leur première présentation, ce qui rendait extrêmement simple leur reconnaissance, et permettait de s'assurer que les participants étaient attentifs à la tâche qu'ils effectuaient. Étant donnée que la base de (Isola et al., 2011b) est, à notre connaissance, la seule disponible pour l'étude

FIGURE 4.8 – Échantillon d'images tirées de la base de données proposée par le MIT (Isola et al., 2011b). Les images sont rangées de la plus mémorable (en haut à gauche) à la moins mémorable (en bas à droite).

de la mémorabilité des images en informatique, il pourrait être intéressant de comparer les scores de mémorabilité obtenus par ces auteurs à des scores obtenus dans des conditions similaires, mais en laboratoire. Notre expérience présentée dans le chapitre 5, mise en place pour obtenir, entre autres, des scores de mémorabilité pour 150 images, a été réalisée en laboratoire, ce qui nous a permis de mettre en place des contrôles difficile à mettre en place en crowdsourcing (p. ex. consignes détaillées dans la langue maternelle des participants, tests de vision et de fatigue, suivi des participants par un expérimentateur qui s'assure de leur bonne compréhension de la tâche). A l'instar de (Isola et al., 2011b), nous avons mesuré la cohérence inter-individuelle des scores de mémorabilité des images (i.e. si les images les plus mémorables pour un groupe d'individus étaient les images les plus mémorables pour un autre groupe d'individus) dans la section 7.1 du chapitre 7 ; nos résultats sont mis en parallèle avec ceux de ces auteurs.

4.2.3 Estimation subjective de la mémorabilité et performance de mémoire

Pour constituer de nouvelles bases de données pour l'étude de la mémorabilité, ou élargir la base existante, il pourrait être intéressant de recourir à l'annotation manuelle des images, en demandant à des personnes d'évaluer a priori la mémorabilité d'images. En effet, obtenir des scores de mémorabilité à l'aide d'un test de mémoire nécessite la mise en place d'une expérience adaptée et un certain nombre de participants pour ob-

tenir un taux de reconnaissance significatif par image. L'annotation manuelle pourrait s'avérer moins coûteuse en participants, et un jugement de mémorabilité peut, contrairement à un test interrogeant la mémoire à long terme, être effectué sans délai, à l'aide d'une simple question de mémoire (comme dans (Libkuman et al., 2007)). Il est nécessaire, cependant, de s'assurer que le test de mémoire utilisé — en l'occurrence, une tâche de reconnaissance —, objectif, et l'auto-évaluation par les participants, subjective, mesurent des *mémorabilités* similaires ; ou de connaître les liens unissant ces deux *mémorabilités* pour inférer l'une de l'autre.

Isola *et al.* se sont intéressés, à travers deux expériences conduites en crowdsourcing, à la corrélation existant entre un jugement subjectif porté sur la mémorabilité d'une image et une performance de mémoire mesurée par une tâche de reconnaissance (Isola et al., 2014). Dans la première expérience, trente participants devaient indiquer s'ils pensaient que les images qui leur étaient présentées étaient ou non mémorables, en répondant à la question suivante : « Est-ce une image mémorable ? » Dans la seconde expérience, indépendante, trente nouveaux participants devaient indiquer, pour chacune des images qui leurs étaient présentées, s'ils pensaient qu'ils auraient ou non reconnu l'image si celle-ci leur avait été présentée un matin et qu'il l'avait revue le soir du même jour. L'analyse des résultats aux deux tâches montrent que les jugements portés par les participants sur la mémorabilité des images ne corrèlent pas avec les scores mémorabilité des images (obtenus dans (Isola et al., 2011b)). Ces résultats suggèrent donc qu'une annotation manuelle des images, en vue de leur attacher des scores de mémorabilité comparables à ceux calculés à partir d'une mesure de performance de mémoire, est difficilement envisageable.

Cependant, au moins deux études de psychologie, réalisées antérieurement à celle de (Isola et al., 2014), et dont ces auteurs ne font pas mention, ont montré que des jugements portés sur la facilité à mémoriser un matériel verbal prédisaient dans une certaine mesure les performances de rappel ou de reconnaissance subséquentes (Underwood, 1966, Jacob and Nelson, 1990). Or, les jugements portés sur la mémorabilité de stimuli seraient basés sur les jugements portés sur la facilité à mémoriser ces stimuli (Libkuman et al., 2007). Dans (Underwood, 1966), les participants devaient prédire la rapidité avec laquelle ils pensaient apprendre des trigrammes (i.e. des mots de trois lettres). Une liste de vingt-sept trigrammes était présentée aux participants, suivie par un test de rappel libre. Puis la liste était présentée de nouveau, suivie par un nouveau test de rappel libre ; et ainsi de suite, jusqu'à ce que les participants aient rappelé d'un seul coup tous les items. Les résultats montrent que les participants ont prédit leur propre apprentissage avec un succès considérable : la corrélation entre le temps estimé par les participants pour apprendre les trigrammes et le temps réel mis pour les apprendre était supérieure à .63. Dans (Jacob and Nelson, 1990), les participants à l'étude devaient émettre un jugement sur leur facilité à mémoriser les vingt items d'une liste (des paires de noms indépendants, par exemple « TABLE-LAC »), sans qu'ils n'aient été informés qu'ils

auraient ensuite à mémoriser ces items. Les vingt items étaient d'abord présentés simultanément ; la tâche consistait à choisir l'item qu'ils pensaient le plus facile à mémoriser. Lorsqu'un item sélectionné, il était ôté de la liste ; la tâche se poursuivait jusqu'à ce que tous les items aient été choisis. Ensuite, les participants apprenaient les items de la liste. Durant cette phase d'apprentissage, les vingt items étaient présentés l'un après l'autre au centre d'un écran durant quelques secondes, dans un ordre aléatoire. Une phase de test advenait ensuite, où le premier mot de la paire de noms (par exemple, « TABLE ») était présenté, et le mot associé (« LAC ») devait être rappelé par le participant. Les items correctement rappelés étaient ôtés de la liste, et les items non rappelés étaient présentés une nouvelle fois ; et ainsi de suite, jusqu'à ce que les participants aient rappelé correctement tous les items. Quatre semaines plus tard, les participants passaient un test de reconnaissance. Les résultats montrent une corrélation significative de .22 entre les jugements portés par les participants sur leur facilité à apprendre les items et leur performance de reconnaissance subséquente de ces items. Ces résultats sont conformes à ceux obtenus par (Underwood, 1966). Ensemble, les résultats de ces deux études peuvent laisser penser que, puisque nous sommes capables de prédire, dans une certaine mesure, la difficulté à mémoriser des items verbaux, nous pourrions être capables de prédire, au moins dans une certaine mesure, la mémorabilité d'images. Il est possible qu'une telle capacité n'ait pas transparu dans les résultats de (Isola et al., 2014) ; en particulier, on peut noter que le contrôle de tâches hautement subjectives, telles que porter des jugements méta-cognitifs, est très difficile en crowdsourcing, et qu'il reste possible que les résultats, obtenus pour un faible nombre de participants, ne soient pas entièrement fiables. Pour dissiper ce doute, il serait intéressant de mener de nouvelles études en laboratoire. Dans la section 7.1 du chapitre 7, nous apportons des éléments de réponse concernant notre capacité à prédire la mémorabilité des images, et mettons en lumière l'influence de l'émotion véhiculée par les images sur cette capacité.

4.3 Méthodes de prédiction de la mémorabilité d'images

La plupart des approches existantes pour prédire la mémorabilité des images reposent sur l'apprentissage automatique de liens entre des caractéristiques intrinsèques d'images, présentées en entrée, et les scores de mémorabilité de ces images *vérité terrain*, présentés en sortie. En utilisant une telle méthode, Isola *et al.* ont, à notre connaissance, les premiers montré que la mémorabilité des images pouvait être inférée computationnellement des caractéristiques de l'image (Isola et al., 2011b). Le cadre de travail introduit par ces auteurs dans l'étude de la mémorabilité des images sera réutilisé par la plupart des études suivantes (p. ex. (Isola et al., 2011a, Khosla et al., 2012b, Kim et al., 2013, Mancas and Le Meur, 2013, Celikkale et al., 2015)). D'autre part, cette étude aura ouvert la voix à des travaux annexes, tels que la modification de la mémorabilité des images (Khosla et al., 2013) et la conception de techniques de visualisation de données

plus efficaces (Borkin et al., 2013).

4.3.1 Cadre général pour l'apprentissage automatique de la mémorabilité d'images

Pour construire un modèle prédictif de la mémorabilité des images, (Isola et al., 2011b) ont adopté une approche classique en vision par ordinateur, précédemment adoptée pour étudier d'autres propriétés subjectives des images (p. ex. (Luo and Tang, 2008,Liu et al., 2010a)). Lors de la phase d'apprentissage, des caractéristiques de bas et haut niveau sont extraites des images, et utilisées conjointement avec les scores de mémorabilité (i.e. la vérité terrain) pour entraîner un séparateur à vaste marge (SVM) pour la régression. Le SVM est ensuite utilisé pour prédire les scores de mémorabilité de nouvelles images. La figure 4.9 donne une illustration de cette méthode. Depuis lors, un certain nombre de modèles ont été proposés pour améliorer les résultats obtenus par ces auteurs (Isola et al., 2011a,Khosla et al., 2012b,Kim et al., 2013,Mancas and Le Meur, 2013,Celikkale et al., 2015), qui s'inscrivent également dans une logique d'apprentissage automatique supervisée. Les images et les scores de mémorabilité utilisés pour l'apprentissage par ces modèles proviennent de la base de (Isola et al., 2011b) ; les caractéristiques utilisées varient cependant selon les auteurs.

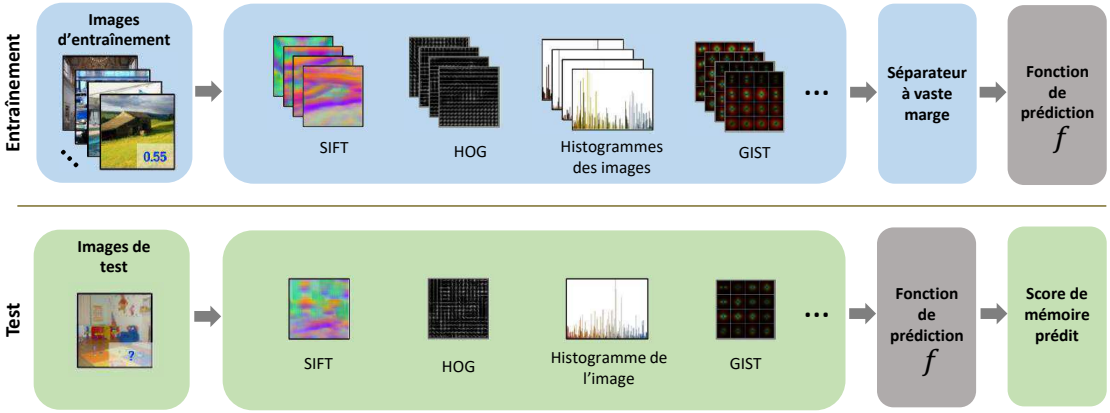


FIGURE 4.9 – Cadre de travail utilisé pour l'apprentissage automatique de la mémorabilité des images (adapté de (Celikkale et al., 2015)).

4.3.2 Perfectionnement des méthodes de prédiction

Plusieurs études se sont également intéressées aux modèles computationnels d'attention visuelle dans un objectif de prédiction de la mémorabilité des images (p. ex. (Mancas and Le Meur, 2013, Celikkale et al., 2015)). Dans (Mancas and Le Meur, 2013),

les auteurs proposent, comme nous l'avons expliqué dans la section 4.1.1, deux nouvelles caractéristiques extraites à l'aide de modèles d'attention visuelle : la couverture de saillance des images et le contraste des structures présentes dans l'image. Ces caractéristiques permettent de remplacer avantageusement un certain nombre de caractéristiques utilisées par (Isola et al., 2011b) pour la prédiction de la mémorabilité des images. Dans (Celikkale et al., 2015), les auteurs s'intéressent aux caractéristiques non plus de l'image dans sa globalité, mais des régions de l'image les plus saillantes, pour prédire la mémorabilité globale des images. L'intérêt pour les modèles d'attention visuelle dans notre champ de recherche nous a conduit à étudier la performance de plusieurs de ces modèles, parmi les plus performants ou récents, en lien avec la mémorabilité des images et l'émotion qu'elles véhiculent. C'est l'objet du chapitre 11.

Plus récemment, (Bylinskii et al., 2015b) ont proposé, comme on l'a évoqué, une méthode pour prendre en compte automatiquement le contexte dans lequel une image est présentée pour en prédire la mémorabilité. Le modèle proposé par ces auteurs calcule automatiquement l'influence de changements du contexte de présentation des images sur leur mémorabilité. En particulier, le modèle permet de calculer le degré de distinctivité d'une image par rapport à son contexte de présentation (i.e. aux autres images vues dans une tâche de reconnaissance similaire à celle utilisée par (Isola et al., 2011b), présentée dans la figure 4.7), sur la base d'un ensemble de caractéristiques extraites des images. Nous reviendrons sur ce point dans le chapitre 9, où nous explorons deux voies différentes pour prendre en compte la distinctivité des images en matière d'information émotionnelle intrinsèque par rapport à leur contexte de présentation pour en prédire la mémorabilité.

La technique de l'apprentissage profond, qui a conduit à des progrès considérables dans un grand nombre de domaines, en particulier en traitement d'images, de vidéos, ou encore de la parole (LeCun et al., 2015), n'a, à notre connaissance, pas encore été essayée pour la prédiction de la mémorabilité des images. Selon Lecun *et al.*, cette technique permet d'entraîner des modèles computationnels composés de multiples couches de traitement, pour apprendre des représentations de données avec de multiples niveaux d'abstraction. Dans le chapitre 8, nous introduisons l'apprentissage profond pour la prédiction de la mémorabilité des images, et comparons les performances de notre modèle avec celles de modèles précédents.

4.4 Conclusion

Dans ce chapitre, nous avons donné un aperçu des études portant sur la mémorabilité des images en informatique. Ces études auront principalement cherché à mettre en lumière des caractéristiques spécifiques aux images mémorables, dans l'objectif de développer un modèle représentatif de ce type d'image. Plusieurs études ont utilisé avec succès des modèles de vision par ordinateur, suscitant l'intérêt pour ces outils dans notre

champ de recherche. D'autre part, une étude a récemment proposé un modèle qui exploite automatiquement le contexte de présentation des images ; à notre connaissance, c'est la première étude qui s'intéresse à des facteurs extrinsèques d'une image pour en prédire la mémorabilité. Les facteurs individuels, à nos yeux de première importance, n'ont cependant, à notre connaissance, pas encore été étudiés dans le cadre de l'étude de la mémorabilité des images en informatique. À l'heure où les données utilisateurs se multiplient, aussi bien que les senseurs qui permettent de collecter en temps réel de telles données, il nous semble que c'est là un point important à développer pour parvenir à une prédiction réellement précise de la mémorabilité des images. D'autre part, l'apprentissage profond n'a pas encore été essayé pour la prédiction de la mémorabilité des images, malgré la multiplication récente des preuves de l'efficacité d'une telle technique en traitement d'images. Finalement, on pourra avoir remarqué que nous n'avons pas fait mention d'études ayant porté sur la prédiction de la mémorabilité de vidéos ; c'est que, à notre connaissance, de telles études n'existent pas. Ce serait là, comme nous le proposerons à la fin de cette thèse, une perspective intéressante pour élargir nos travaux.

Conclusion

Cette première partie était consacrée au rapprochement théorique entre la mémorabilité des images telle qu'elle est étudiée en informatique et la mémoire humaine telle qu'est est étudiée en psychologie. Nous avons fait apparaître que la mémorabilité étudiée en vision par ordinateur n'est qu'une des « mémorabilités » possibles. En effet, la mémoire est multiple, comme le sont ses manifestations et les manières de la mesurer. Elle est également sujette à de nombreuses influences. En particulier, l'émotion véhiculée par les images, leur contexte de présentation, l'observateur qui les regarde et la manière dont il les regarde, et la durée de leur stockage mnésique, sont des facteurs qui, lors d'une tâche de reconnaissance, sont susceptibles d'influencer la performance de mémoire mesurée. L'intégration de ces facteurs dans les modèles de prédiction de la mémorabilité des images pourrait, par conséquent, conduire à des prédictions plus précises. D'autre part, la mise en place de nouvelles techniques pour la prédiction de la mémorabilité des images, comme l'apprentissage profond et l'intégration computationnelle de facteurs extrinsèques des images, pourrait substantiellement améliorer la précision des modèles.

La suite de cette thèse est consacrée à la présentation des différents travaux que nous avons réalisés, qui s'inscrivent dans une volonté de mettre à profit ces différents leviers pour améliorer notre capacité à prédire la mémorabilité des images. Elle commence par la présentation d'une nouvelle base de données, qui vient s'ajouter à la seule base actuellement disponible, à notre connaissance, pour l'étude de la mémorabilité des images. Ce matériel, que nous rendrons disponible pour la communauté scientifique, constitue le point de départ de nos contributions.

II

Des scores d'émotion et de mémorabilité pour des images numériques

Introduction

À notre connaissance, une seule base de données est disponible pour l'étude de la mémorabilité des images en vision par ordinateur. Les images qu'elle contient sont associées chacune à un score de mémorabilité, qui correspond à une performance de reconnaissance moyenne de l'image mesurée quelques minutes après son encodage. Les études portant sur la prédiction computationnelle de la mémorabilité des images ont largement reposé sur l'utilisation de cette base, en particulier pour l'entraînement et l'évaluation des modèles. La création d'une nouvelle base de données serait profitable, pour évaluer la capacité des modèles de prédiction à généraliser. D'autre part, la création d'une telle base nous permettrait d'obtenir du matériel pour étudier les différents facteurs susceptibles d'influencer la mémorabilité des images, sur lesquels la première partie de cette thèse a jeté de la lumière, ou de les contrôler.

Dans le chapitre 5, nous présentons une nouvelle base de donnée pour l'étude de la mémorabilité des images. Cette base est constituée de 150 images, associées chacune à deux scores de mémorabilité, correspondant à des performances moyennes de mémoire à long terme mesurées quelques minutes après l'encodage des images et un jour après, respectivement. Chaque image a également été évaluée sur les dimensions d'arousal et de valence. En outre, nous avons enregistré les données oculométriques des observateurs des images, et recueilli un certain nombre de données individuelles, à l'aide de diverses questions et questionnaires de psychologie, pour mieux connaître nos participants.

Le chapitre 6 traite de l'analyse des scores d'émotion obtenus dans le chapitre précédent, et de leur mise en relation avec les scores obtenus dans les études précédentes.

Le chapitre 7 porte sur l'analyse conjointe des scores d'émotion et de mémorabilité. Nous nous y intéressons aux effets de l'émotion sur la mémorabilité des images, et investiguons la question du rôle de l'émotion dans l'évolution en mémoire à long terme de la mémorabilité des images.

5

Une nouvelle base de données pour l'étude de la mémorabilité des images

Les études qui, à ce jour, ont porté sur la mémorabilité des images en vision par ordinateur ont toutes, à notre connaissance, soit exploité la base de (Isola et al., 2011b), soit obtenus des scores de mémorabilité en utilisant une méthode similaire à celle employée par ces auteurs (Isola et al., 2011a, Khosla et al., 2012a, Khosla et al., 2012b, Mancas and Le Meur, 2013, Khosla et al., 2013, Celikkale et al., 2013, Kim et al., 2013, Isola et al., 2014, Celikkale et al., 2015, Bylinskii et al., 2015b). Cette unique vérité terrain pose problème : outre la reproduction de biais potentiels qu'une unique base de donnée est susceptible de causer dans les travaux qui l'utilisent, cette absence de diversité ne nous permet pas d'évaluer la capacité des modèles de prédiction à généraliser. D'autre part, notre étude de la littérature a suscité notre intérêt pour un certain nombre de facteurs susceptibles d'influencer la mémorabilité des images. L'objectif de ce chapitre est de présenter une nouvelle base de données qui nous permette d'étudier ces différents facteurs.

5.1 Les apports de cette nouvelle base de données

A la lumière des connaissances présentées dans la première partie de cette thèse, nous avons identifié plusieurs points qu'il nous est apparu important de prendre en compte pour la création d'une nouvelle base de données pour l'étude de la mémorabilité des images. Ces points concernent l'émotion véhiculée par les images, la durée de rétention

mnésique séparant l'encodage des images de leur récupération, les données extrinsèques — contextuelles et individuelles — des images, et l'attention visuelle des observateurs des images.

5.1.1 Des scores d'émotion pour l'étude de la mémorabilité

Un des premiers points dont nous avons mesuré l'importance en étudiant la littérature concerne l'émotion véhiculée par les images. En effet, comme nous l'avons établi dans le chapitre 3, émotions et mémoire sont étroitement liées. La répartition des images d'une base de données destinée à l'étude de la mémorabilité dans l'espace émotionnel est susceptible d'entraîner des biais dans les modèles entraînés à partir des images constituant cette base, comme nous le montrons dans la section 8.4 du chapitre 8. D'autre part, la théorie de la consolidation des souvenirs suggère que la consolidation pourrait ne pas s'appliquer uniformément à l'ensemble des souvenirs d'images, en raison de l'émotion qu'elles véhiculent (McGaugh, 2000). Plus précisément, la mémoire des stimuli émotionnellement chargés tendrait à être mieux préservée que la mémoire des stimuli neutres (LaBar and Phelps, 1998). Par conséquent, il est possible que le passage du temps bouleverse l'ordre de mémorabilité des images mémorisées, en raison de l'émotion qu'elles véhiculent. C'est un problème si l'on considère que les scores de mémorabilité des images actuellement disponibles pour notre communauté renvoient à un phénomène de mémoire immuable, ou même durable. Il semble donc important de prendre en compte l'émotion véhiculée par les images dans l'étude de leur mémorabilité.

Dans un cadre de prédiction automatique de la mémorabilité, on s'intéressera en particulier aux études, assez récentes, qui ont porté sur l'extraction computationnelle de l'information émotionnelle des images (p. ex. (Wang and Yu, 2005, Kim et al., 2005, Joshi et al., 2011)). L'information émotionnelle n'a, à notre connaissance, encore jamais été utilisée dans un objectif de prédiction de la mémorabilité d'images. Dans un tel objectif, il serait intéressant, pour étudier les liens entre émotion et mémorabilité, de pouvoir disposer d'une base de données d'images associées à des scores sur ces différentes dimensions. Ce n'est pas, à notre connaissance, le cas de la base de (Isola et al., 2011b). Nous espérons que la création d'une telle base encouragera le rapprochement entre les études portant sur la prédiction de la mémorabilité, et celles portant sur la prédiction de l'émotion véhiculée par les images.

Deux jeux de données peuvent être rapprochés d'une telle base de données. Grühn et Scheibe (Grühn and Scheibe, 2008) ont rapporté des scores de mémorabilité pour 504 images de l'IAPS (Lang et al., 1997). Les participants à cette étude passaient une tâche de reconnaissance (quelques minutes après l'encodage initial des images), à partir des résultats de laquelle les auteurs ont calculé des scores de mémorabilité. Libkuman *et al.* ont également rapporté des scores de mémorabilité pour les images de l'IAPS (Libkuman et al., 2007). Ces scores correspondent cependant à un jugement porté a

priori sur la mémorabilité des images, et non à une mesure objective d'une performance de mémoire, telle qu'on peut en obtenir en utilisant une tâche de reconnaissance. Les scores de mémorabilité obtenus dans ces études ne correspondent pas à la définition de la mémorabilité donnée par (Isola et al., 2011b). Il pourrait toutefois être intéressant de les considérer (ce que nous faisons dans le chapitre 7, où nous comparons les scores de mémorabilité et d'émotion de la base de données créée à partir de l'expérience présentée dans ce chapitre avec ceux obtenus par ces auteurs). À part ces deux études, il n'existe pas, à notre connaissance, de collection disponible d'images évaluées à la fois sur les dimensions d'émotion et de mémorabilité.

5.1.2 Une double mesure de la performance de mémoire à long terme

Comme nous l'avons précédemment souligné, une des limites de la base de données de (Isola et al., 2011b) porte sur la durée de la rétention mnésique entre l'encodage des images et leur récupération, déterminée par la tâche de reconnaissance utilisée par ces auteurs pour obtenir des scores de mémorabilité. En effet, les performances de reconnaissance des images cibles ont été mesurées quelques minutes après la première présentation des images. Si, comme nous l'avons souligné, c'est assez pour considérer que la mémorabilité reflète une performance moyenne de mémoire à long terme, cela ne permet pas de déterminer si cette mémorabilité est pérenne. En particulier, selon la théorie de la consolidation, certains souvenirs sont consolidés et d'autres non (McGaugh, 2000). En testant la mémoire à deux moments différents en mémoire à long terme, nous pourrions déterminer l'influence de la consolidation des souvenirs sur la mémoire des images.

Il serait intéressant de séparer la réalisation de ces deux mesures par un intervalle de rétention durant lequel l'oubli est significatif. Pour ce faire, nous avons précédemment proposé de nous aider de la courbe d'oubli en mémoire à long terme d'Ebbinghaus (Ebbinghaus, 1913). Cette courbe (voir la figure 1.3) montre une diminution très rapide de la performance de mémoire durant la première journée suivant la phase d'encodage, après quoi la performance tend à se stabiliser. En somme, il serait intéressant de mesurer la mémoire une première fois quelques minutes après l'encodage des images, tel que cela a été fait par (Isola et al., 2011b), et une seconde fois un jour après. Le sommeil jouant un rôle important dans la consolidation (Stickgold, 2005), un tel procédé aurait l'avantage de prendre en compte ce facteur.

Ces deux mesures de la performance de mémoire nous permettraient d'obtenir deux scores de mémorabilité pour chaque image de notre base de données, calculés à partir des résultats aux tâches reconnaissance réalisées quelques minutes et un jour après l'encodage des images, respectivement. De tels scores nous permettraient d'évaluer dans quelle mesure des images mémorables quelques minutes après leur encodage le sont encore après une durée de rétention d'un jour. S'il s'avérait que la baisse de mémorabilité des images est significative durant cet intervalle de rétention, il pourrait être intéressant

de chercher à lier les caractéristiques de bas et haut niveau des images avec la rapidité de l'oubli en mémoire à long terme. On pourrait alors recourir à une méthode d'apprentissage automatique, telle que celle présentée dans la figure 4.9 du chapitre 4, et commencer par étudier les caractéristiques déjà utilisées pour la prédiction de la mémorabilité, ainsi que l'émotion véhiculée par les images. On chercherait alors à répondre à la question suivante, qui, à notre connaissance, n'a encore jamais été posée en vision par ordinateur : « Quelles sont les caractéristiques d'une image qui font que sa mémorabilité se maintient dans le temps ? »Nous développons cette question dans les perspectives de cette thèse, présentées dans la conclusion générale 12.5 ; en particulier, nous y envisageons d'augmenter la taille de notre base de données, afin qu'elle atteigne une échelle suffisante pour qu'il soit possible d'utiliser les techniques d'apprentissage automatique dans des conditions satisfaisantes.

5.1.3 Des données oculométriques pour les images

Comme nous l'avons expliqué dans la section 4.3.2 du chapitre 4, les modèles computationnels d'attention visuel suscitent l'intérêt des chercheurs travaillant sur la prédiction de la mémorabilité des images (Khosla et al., 2012b, Mancas and Le Meur, 2013, Celikkale et al., 2015). Et il est probable que l'engouement pour ces modèles dans notre champ de recherche se poursuive. Cependant, le comportement des modèles d'attention visuelle pour des images dont la coloration émotionnelle ou le degré de mémorabilité varie n'est pas bien connu. En vision par ordinateur, les modèles d'attention visuelle sont principalement basés sur le concept de cartes de saillance (Riche et al., 2013). Typiquement, un modèle permet de calculer une carte de saillance à partir d'une image, qui prédit les endroits de cette image où un observateur humain est le plus susceptible de porter son attention. Il serait intéressant d'obtenir des cartes de saillance *vérité terrain* pour les images de notre base de données, pour pouvoir estimer la performance de différents modèles d'attention visuelle, en lien avec les scores d'émotion et de mémorabilité des images. De telles cartes de saillance peuvent être calculées à partir des données oculométriques des observateurs des images (Le Meur and Baccino, 2013). D'autre part, il serait intéressant de mieux comprendre comment l'attention visuelle est liée à la mémorabilité. Mancas et Le Meur ont entrepris une telle étude (Mancas and Le Meur, 2013). Ils ont montré que la durée des fixations oculaires tendait à diminuer avec le degré de mémorabilité des images. Cependant, ces résultats méritent d'être confirmés, d'autant que les effets trouvés par les auteurs sont ténus. L'émotion est également liée à l'attention visuelle. En particulier, nous avons vu dans la section 3.1.1 du chapitre 3 que l'émotion suscitée par un stimulus est susceptible d'influencer l'encodage mnésique. Enregistrer les données oculométriques des observateurs des images de notre base de données nous permettrait d'étudier les liens entre attention visuelle, émotions et mémorabilité. Dans cet objectif, nous pourrions utiliser un oculomètre. Cet outil, présenté dans la section 2.3.2 du chapitre 2, nous permettrait d'enregistrer les mouvements ocu-

laires des participants à notre étude, et, à partir de cette mesure, d'étudier l'attention visuelle en lien avec les facteurs qui nous intéressent.

5.1.4 Des données liées aux observateurs des images

Les facteurs individuels font figure de grands absents dans les travaux portant sur la prédiction de la mémorabilité des images. Pourtant, comme nous l'avons expliqué dans la section 4.1.2 du chapitre 4, la prise en compte de ces facteurs dans les modèles de prédiction pourrait nous conduire à des prédictions plus précises de la mémorabilité. Dans cet objectif, il serait intéressant de modéliser les liens entre facteurs individuels et mémorabilité des images. Aussi, il nous parait intéressant de collecter des données sur les participants de notre étude, à l'aide de question et questionnaires de psychologie, en vue d'en établir des profils. Le détails de ces données, aussi bien que les scores de mémorabilité et d'émotion *moyens* pour les images, seront rendus disponibles à la communauté scientifique. Cela pourrait encourager d'autres auteurs à faire de même ; en effet, il est encore difficile, aujourd'hui, d'avoir accès aux données individuelles à partir desquelles les différentes informations *moyennes* associées aux contenus numériques des bases de données disponibles ont été calculées. Cette réalité est probablement en partie responsable du faible nombre de modèles intégrant les idiosyncrasies dans leurs équations.

Objectif de cette étude

Pour résumer, les études existantes qui portent sur la prédiction de la mémorabilité ont, à notre connaissance, laissé de côté les facteurs suivants, susceptibles d'influencer la mémorabilité des images : 1/ les émotions véhiculées par les images, 2/ la durée de rétention des images en mémoire à long terme (durant laquelle interviennent les processus de consolidation et d'oubli) et 3/ les facteurs individuels. D'autre part, nous sommes également intéressés par l'attention visuelle des observateurs des images. L'étude présentée dans ce chapitre a pour objectif de constituer une nouvelle base de données pour l'étude de la mémorabilité des images, qui permettent d'étudier les liens entre ces différents facteurs et la mémorabilité des images. Elle vise à obtenir, pour chacune des 150 images qui constitueront la base : deux scores de mémorabilité correspondant à une mesure de la mémoire effectuée soit quelques minutes après l'encodage, soit 24h après, respectivement ; un score d'arousal et un score de valence, en accord avec l'approche dimensionnelle des émotions adoptée dans cette thèse ; les données oculométriques des participants ; des informations sur les participants, collectées au moyen de questions et de questionnaires psychologiques. Nous désirons également obtenir des scores de mémorabilité comparables à ceux proposés par (Isola et al., 2011b). Pour cette raison, notre méthode de mesure de la mémoire des images est proche de celle employée par ces auteurs. D'autre part, afin d'obtenir des images véhiculant des émotions variées, nous

avons utilisé des images de l'IAPS (Lang et al., 2008), qui ont déjà été évaluées sur les dimensions d'arousal et de valence. La comparaison de nos scores d'émotion avec ceux des précédents auteurs qui ont rapporté de tels scores pour des images de l'IAPS (Grühn and Scheibe, 2008,Ito et al., 1998,Lang et al., 2008,Libkuman et al., 2007) nous permettra également de nous assurer de la cohérence de nos données.

Notre protocole expérimental est décrit dans la suite de ce chapitre.

5.2 Matériel et méthode

Participants

Cinquante participants (de 18 à 41 ans ; $\bar{x} = 22.54; SD = 5.01$; dont 60% de femmes), rémunérés pour leur participation, ont été recrutés à Nantes. A leur arrivée sur les lieux de la passation, les participants ont été ventilés entre trois groupes différents, suivant notre plan d'expérience (voir Table 5.1).

Matériels et outils employés

Stimuli

Les 625 images utilisées dans l'étude ont été sélectionnées aléatoirement dans l'IAPS (Lang et al., 2008), après que nous ayons retiré les items qui sont des duplications modifiées de photographies originales. Par exemple, l'image 6570.2, représentée dans la figure 5.1 est dérivée de l'image originale 6570.1, modifiée de telle sorte que l'arme de poing est devenue un sèche-cheveux. Les images étaient ensuite réparties aléatoirement dans six groupes différents, suivant notre plan expérimental illustré par la table 5.1. Trois des groupes étaient composés de 50 images « cibles », présentées deux fois dans les tâches de mémoire : une fois pour leur mémorisation, et une seconde fois pour mesurer la performance de reconnaissance des participants. Les trois autres groupes étaient composés de 200 images de « remplissage », qui apparaissaient une seule fois pour chaque participant.

Toutes les images étaient vues un nombre de fois similaire : le plan d'expérience spécifique a été conçu pour optimiser le nombre de scores collectés tout en limitant un possible biais contextuel, susceptible d'apparaître même si les images sont présentées aléatoirement, puisqu'une image peut être plus saillante dans un jeu d'images que dans un autre (Bylinskii et al., 2015b). Nous avons contrôlé qu'aucune des catégories d'images n'était sur-représentée dans un groupe d'images (le critère étant le suivant : pas plus de trois images de la même catégorie — p. ex. trois serpents — par test, c'est-à-dire sur 200 ou 250 images) ; nous avons remplacé les images lorsque ça a été nécessaire.

Nous avons également testé l'équivalence des groupes, en nous basant sur les distributions intra-groupes des scores d'arousal et de valence. Nous avons effectué un test de

(a) Image 6570.1 (b) Image 6570.2

FIGURE 5.1 – Une image originale de l'IAPS et sa version modifiée.

Kruskal-Wallis pour tester l'hypothèse nulle selon laquelle les scores d'arousal des différents groupes d'images suivaient une distribution similaire ($\chi^2(5) = 5.15, p = .398$), et un autre pour les scores de valence ($\chi^2(5) = 4.58, p = .469$). Comme les hypothèses nulles ne peuvent être rejetées, les résultats signifient que les scores d'arousal et de valence dans nos six groupes d'images se distribuent de manière équivalente. En positionnant les images dans un espace cartésien selon leurs scores de valence (portée en abscisse) et d'arousal (porté en ordonnée), nous avons observé pour les six groupes que les images se répartissaient selon une fonction en U, similaire à celle observée par Lang *et al.* pour l'ensemble des images de l'IAPS (Lang et al., 1997).

Les 25 images qui ne faisaient partie d'aucun groupe étaient utilisées dans les phases d'entraînement, pour familiariser les participants avec les tâches à effectuer (les données collectées pour ces images ne sont pas prises en compte dans les analyses).

Tests de mémoire

L'expérience incluait deux tests de mémoire. Le premier test de mémoire consistait en une phase d'encodage mnésique et une phase de test entrelacées. Durant la tâche (d'une durée approximative de 15 minutes), les participants voyaient une séquence d'images, chacune d'entre elle affichée pour deux secondes, séparées un écran noir affiché une seconde (voir la Figure 5.2(a)). La consigne relative à la tâche était de presser la barre d'espace dès qu'une image déjà vue réapparaissait. Une session de test était composé de 50 images cibles (i.e. les images répétées une fois) et de 200 images de remplissage. Le rôle des images de remplissage était double : d'une part, elles permettaient d'espacer la première apparition d'une cible de sa répétition ; d'autre part, elles incluaient les cibles du second test de mémoire (effectué le lendemain), de sorte que la performance de mémoire ne soit pas mesurée sur des images déjà répétées. Les images étaient présentées

TABLE 5.1 – Plan d'expérience

Groupe	Test 1	Test 2	Notation
1	T1[a], F1(⊂T2)	T2, F2	T1, T2
2	T2[a], F2(⊂T3)	T3, F3	T2, T3
3	T3[a], F3(⊂T1)	T1, F1	T3, T1

Note — Le test 1 est composé de 50 images cibles (T) répétées une fois et de 200 images de remplissage (F) non répétées. Le test 2 est composé de 50 images cibles vues comme images de remplissage dans le test 1, et de 200 nouvelles images de remplissage. La notation des images sur les dimensions d'arousal et de valence porte sur les 100 images cibles préalablement vues.

dans un ordre pseudo-aléatoire, avec la restriction suivante : la répétition d'une cible ne pouvait se produire que si un espacement d'au moins 70 images (3'30 min) la séparait de sa première apparition, de sorte que le test de reconnaissance mesurait une performance de MLT. Lorsque la barre d'espace était pressée, l'image était encadrée par un rectangle bleu pour informer le participant de la prise en compte de sa réponse.

Le second test de mémoire était similaire au premier (voir Figure 5.2(b)). La seule différence était que ne lui était pas adjoint une phase d'encodage mnésique, puisque les cibles étaient des images de remplissage vues lors de la précédente tâche (si bien que la tâche était plus courte, durant approximativement 12'30 min).

Les deux tâches reconnaissance commençaient avec un rappel succinct de la consigne (préalablement explicitée par l'expérimentateur) suivi d'une phase d'entraînement.

Tâche d'évaluation de l'émotion induite par les images

La tâche portait sur les images cibles vues auparavant dans les tâches de reconnaissance (voir Figure 5.2(c)). Un écran invitant le participant à évaluer la prochaine image était affiché pendant deux secondes ; ensuite, l'image était affichée durant six secondes ; finalement, le participant évaluait l'image, en prenant le temps qu'il désirait pour son évaluation. Une version informatique des échelles SAM, à neuf degrés, était utilisée pour évaluer l'image sur les dimensions d'arousal et de valence (voir Figure 2.4). Avant de passer à l'image suivante, les participants devaient confirmer leur évaluation afin d'éviter les erreurs de manipulation. Les images étaient affichées aléatoirement, pour une durée totale d'environ 30 minutes. La tâche commençait avec un rappel de la consigne

(qui insistait de nouveau sur la définition d'arousal et de valence), suivi d'une phase d'entraînement composée d'images représentant différentes parties de l'espace émotionnel arousal-valence, de sorte que les participants soient familiers à la fois avec la tâche à effectuer et avec le système de notation.

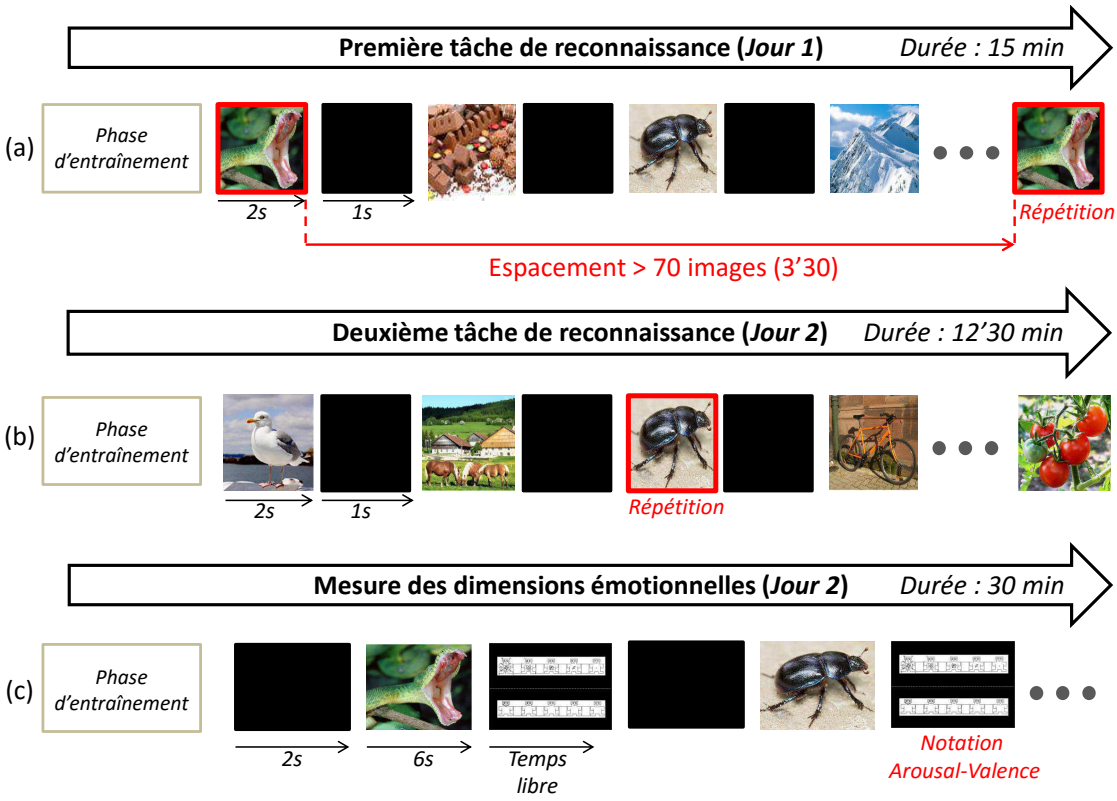


FIGURE 5.2 – Les trois tâches expérimentales. (a) Les tâches d'encodage et de reconnaissance entrelacées le premier jour. (b) La seconde tâche de reconnaissance 24 heures plus tard (les cibles sont des images de remplissage du jour précédent). (c) La tâche de notation des images sur les dimensions d'arousal et de valence avec les échelles SAM. Pour les trois tâches, les images étaient affichées selon un ordre aléatoire.

Questionnaires

Dans la littérature de psychologie, plusieurs facteurs individuels ont été isolés pour leur influence sur l'expérience émotionnelle induite chez l'observateur par une image, sur la mémorabilité d'une image, ou sur ces deux phénomènes, incluant l'âge (Grühn and Scheibe, 2008), le genre (Wrase et al., 2003), l'état affectif à la fois durable (Watkins et al., 1992) et passager (Hermans et al., 1994), et la fatigue (Hart et al., 1987). Pour

collecter des données sur les participants qui nous permettent d'étudier ces facteurs, nous avons utilisé un questionnaire d'information général (âge, genre, etc.) ainsi que plusieurs outils de psychologie.

Le Bem Sex-Role Inventory (BSRI) (Bem, 1981) était utilisé pour mesurer la masculinité-féminité des participants. Ce questionnaire auto-administré est composé de 60 items, pour chacun desquels la personne qui s'évalue doit noter à l'aide d'une échelle en sept points (1 : " jamais ou presque jamais vrai ", 7 : " toujours ou presque toujours vrai ") dans quelle mesure il est adapté pour décrire sa personnalité. La mesure est basée sur des descripteurs qui font référence à différents traits et comportements que les hommes et les femmes considèrent comme valorisants pour les hommes, et pour les femmes (i.e. les caractéristiques de genre socialement valorisés). Le questionnaire inclut 20 descripteurs " féminins " (p. ex. doux, compatissant), 20 descripteurs " masculins " (p. ex. indépendant, agit comme un leader) et 20 traits considéré comme neutre (p. ex. sincère, flexible). Sur la base de ses réponses, le BSRI permet de catégoriser la personnalité de l'individu comme étant globalement " masculine ", " féminine ", ou " indifférenciée " (i.e. que les pôles masculin et féminin sont sous-investis).

L'International Positive and Negative Affect Schedule, Short Form (I-PANAS-SF) (Thompson, 2007) était utilisé pour mesuré l'état affectif dominant du participant au cours de la dernière année. C'est une version raccourcie du PANAS (Watson et al., 1988) qui ne contient que 10 items, destinée à une utilisation internationale avec de utilisateurs dont l'anglais n'est pas la langue maternelle. Les participants qui complétaient l'I-PANAS-SF devaient évaluer dans quelle mesure ils étaient en accord avec chacun des items à l'aide d'une échelle type Likert en sept points (1 : " très légèrement/jamais " ; 7 : " beaucoup/souvent "). L'I-PANAS-SF fournit des résultats sur deux dimensions : la positivité de l'état affectif, et sa négativité.

La matrice de l'humeur (*mood matrix* ; Figure 5.3) proposée par Eich et Metcalfe (Eich and Metcalfe, 1989) était utilisée pour évaluer l'humeur des participants au début et à la fin de chacune des tâches (les deux tests de mémoire et la tâche de notation). Cette matrice est composée de (9x9) 81 cases ; elles mesure simultanément l'arousal (axe vertical) et la valence (axe horizontal). Le centre de la matrice représenterait une humeur neutre. Les participants devaient mettre une croix dans la case qui reflétait le mieux leur humeur du moment (p. ex. un participant qui aurait fait l'expérience d'une humeur positive et stimulante aurait dû barrer d'une croix une case en haut à droite de la matrice).

Enfin, pour mesurer le niveau de fatigue des participants, nous avons utilisé un unique item, extrait du PANAS (à savoir : " Indiquez dans quelle mesure vous vous sentez fatigué. " 1 " très fatigué " ; 5 : " absolument reposé ").

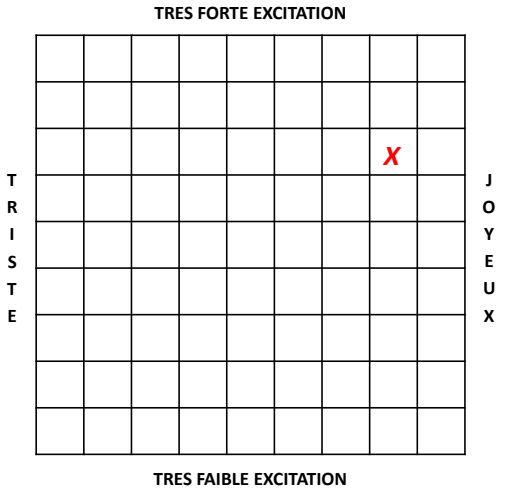


FIGURE 5.3 – La matrice de l'humeur (adaptée de (Eich and Metcalfe, 1989)).

Installations et appareils

L'expérience s'est déroulée dans les locaux de l'IRCCyN (Nantes, France), dans une pièce conçue et équipée pour l'évaluation de la qualité d'images. Les images étaient affichées sur un moniteur de 40 pouces (TV-LOGIC LVM401), avec une résolution d'affichage de 1920 $\times$ 1080. Les participants étaient assis à une distance de 150 centimètres de l'écran (trois fois la hauteur de l'écran). Les images (1024 $\times$ 768 px) étaient centrées sur un fond noir ; à la distance de 150 cm, les images représentaient 18,85 degrés d'angle visuel vertical (il faut noter que ce calcul ne prend pas en compte les bandes noires présentes sur certaines images de l'IAPS). L'expérience a été écrite en Matlab, et avait recours à la Psychtoolbox-3 (Brainard, 1997, Kleiner et al., 2007). Pendant les tests de mémoire et la tâche de notation, un oculomètre SMI RED enregistrait les mouvements des yeux des participants. Le système comprend une caméra infrarouge et deux sources de lumière infrarouge, une de chaque côté de la caméra. L'oculomètre enregistre des points de regard (PG) à une fréquence de 50 PG/sec pour chaque œil. Une calibration du système pour chaque utilisateur était réalisée au début de chaque phase de test, en utilisant 5 points de calibration affichés sur l'écran.

Procédure

À leur arrivée pour la première session de test, les participants recevaient une copie des instructions et il leur était demandé de lire et de signer le formulaire de consentement s'ils souhaitaient participer à l'expérience. Ensuite, ils passaient un test d'acuité visuel (les échelles Monoyer) et un test de vision des couleurs (test de Ishihara). Les participants étaient ensuite conduits dans la salle d'expérience. Ils complétaient une version

numérisée du questionnaire d'information général, une du BSRI, et une du I-PANAS-SF. Ensuite, ils répondaient à la question portant sur leur niveau de fatigue et remplissaient la matrice de l'humeur (action qu'ils étaient amenés à répéter au début et à la fin de chaque session de test). La première session de test était ensuite lancée, correspondant au premier test de mémoire. Le lendemain, après environ 24 heures (+- 3 heures), les participants revenaient pour passer le second test de mémoire. Après une pause de 10 minutes suivant le test de mémoire, ils effectuaient finalement la tâche de notation des images sur les dimensions d'arousal et de valence.

5.3 Calcul des scores d'émotion et de mémorabilité

Un score de mémorabilité pour un image était calculé comme le nombre de détections correctes de ses répétitions sur le nombre total de ses répétitions. Un score de mémorabilité correspond donc à la probabilité qu'une image soit reconnue lorsqu'elle est répétée (une fois) dans un flux d'images. Chaque image a deux scores de mémorabilité : un score correspondant au premier test de mémoire (où la mesure était effectuée quelques minutes après l'encodage) et un score correspondant au second test de mémoire (effectué un jour plus tard). Un score de mémorabilité pour un test de mémoire correspond à un taux de reconnaissance calculée sur 16,6 participants en moyenne (i.e. 50 participants divisés par 3 groupes — voir la table 5.1). Les scores de mémorabilité n'incluent pas d'information sur les fausses alarmes (FA). Une FA advenait lorsqu'un participant pressait la barre espace pour une image qu'il voyait pour la première fois. Le taux moyen de FA était de 0.89% ($\sigma = 2.68\%$) pour la première tâche de reconnaissance, et de $3,69\%$ ($\sigma = 9.73\%$) pour la seconde. Seulement 26 images cibles (sur les 150) pour le premier test de mémoire, et 38 images cibles (sur 150 également) pour le second test, ont donné lieu à au moins une FA lorsqu'elles étaient présentées en condition de remplissage ; la majorité des images notées n'ont donc jamais été reconnues erronément. Comme le taux FA était bas en comparaison du taux de détection correctes, les détections correctes ont peu de chance d'être des confusions chanceuses ; plus précisément, les confusions chanceuses ne devraient pas compter pour plus de 0.89% de l'ensemble des détections correctes dans le premier test de mémoire, et 3.69% dans le second.

Deux scores émotionnels étaient également calculés pour une image : le score d'arousal correspond à la moyenne des valeurs collectées grâce à l'échelle SAM correspondante pour les 33 (ou 34) participants qui ont évalué l'image ; le score de valence était calculé de la même façon.

La figure 5.4 présente un échantillon des images de notre base de données, avec leurs scores d'émotion et de mémorabilité.

FIGURE 5.4 – Les images (a) les moins mémorables et (b) les plus mémorables un jour après leur encodage, avec leurs scores d'arousal, de valence, et de mémorabilité.

Conclusion

Dans ce chapitre, nous avons décrit une expérience qui nous a permis de constituer une nouvelle base de données pour l'étude de la mémorabilité des images, qui vient s'ajouter à la seule base de données, à notre connaissance, disponible actuellement. Le tableau 5.2 résume les différences fondamentales entre ces deux bases de données : la nôtre et celle présentée dans (Isola et al., 2011b).

	Notre base	Base d'Isola *et al.*
Passation	Laboratoire	Crowdsourcing
Nombre d'images	150	2222
Scores de mémorabilité court terme	✓	✓
Scores de mémorabilité long terme	✓	
Scores d'émotion	✓	
Données oculométriques	✓	
Données individuelles	✓	

TABLE 5.2 – Vue d'ensemble des différences entre les deux bases de données destinées à l'étude de la mémorabilité des images actuellement disponibles.

La multiplication des images associées à des scores de mémorabilité, en particulier lorsque les images proviennent de différentes sources, est intéressantes pour la mise au point des modèles prédictifs et l'évaluation de leur capacité à généraliser. Cette base nous aura également permis d'étudier plusieurs facteurs que notre étude de la littérature a permis de distinguer pour leur influence potentielle sur la mémorabilité des images. Dans le chapitre 7, nous étudions les liens entre les émotions véhiculées par les images et leur mémorabilité, en considérant l'influence de l'émotion sur la baisse de mémorabilité des images en mémoire à long terme. Dans le chapitre 10, nous proposons un modèle des liens entre les facteurs individuels mesurés dans cette expérience et la probabilité de reconnaître une image. Dans le chapitre 8, nous montrons la difficulté de notre modèle à apprentissage profond, entraîné sur la base de (Isola et al., 2011b), à généraliser sur notre base de données, et montrons également que notre modèle présente un biais de prédiction en faveur des images véhiculant des émotions négatives et activatrices. Dans le chapitre 11, à partir des données oculométriques des participants à cette expérience, nous étudions le lien entre attention visuelle, émotion et mémorabilité, et évaluons la performance de plusieurs modèles computationnels d'attention visuelle pour les 150 images de notre base de données.

Le chapitre suivant traite de l'analyse des scores d'émotion des images de notre base de données, et de leur mise en relation avec les scores obtenus dans les études précédentes.

6

Généralisabilité des scores d'émotion pour des images

L'IAPS est très utilisée. Plusieurs milliers d'auteurs ont cité les articles présentant cette base de données (Lang et al., 1997, Lang et al., 1999, Lang et al., 2008). Certains de ces auteurs ont fait évaluer à nouveau tout ou partie des images de la base sur les dimensions d'arousal et de valence, et ont rendu accessibles les scores obtenus à la communauté scientifique (Grühn and Scheibe, 2008, Ito et al., 1998, Libkuman et al., 2007). Dans ce chapitre, nous comparons nos scores d'émotion avec ceux obtenus dans ces études, l'objectif étant de déterminer à quel point nos résultats sont généralisables. En particulier, nous avons fait évaluer par une population française, en 2015, des images initialement évaluées par une population américaine dans les années 1990 (Lang et al., 1997). Or l'époque d'évaluation (Libkuman et al., 2007) et l'origine des évaluateurs (Ribeiro et al., 2005) semblent avoir une influence sur l'évaluation des images. Nous nous intéressons également à la forme géométrique de la relation entre les dimensions d'arousal et de valence, qui, ainsi que nous l'avons évoqué dans la section 2.1.3 du chapitre 2, porte à débat, en particulier lorsque les émotions sont suscitées par des images. Nous quantifions également le degré de généralisabilité de l'émotion véhiculée par une image, pour répondre à la question suivante : « Dans quelle mesure une image induit-elle une émotion similaire chez des personnes différentes ? »

6.1 Scores d'émotion pour des images : les différences inter-études

Dans l'expérience décrite au chapitre 5, nous avons obtenus des scores d'émotion pour 150 images de l'IAPS. La notation a été effectuée à l'aide des échelles SAM d'arousal et de valence, comme dans (Lang et al., 1997, Ito et al., 1998, Lang et al., 2008, Grühn and Scheibe, 2008), et d'une manière proche de celle employée par Libkuman *et al.*, qui ont utilisé des échelles non graphiques, en neuf points également (Libkuman et al., 2007). Ceci nous a permis de comparer nos scores d'émotion à ceux de ces auteurs (voir Table 6.1). Alors que les scores d'émotion de Lang *et al.* et de Libkuman *et al.* concernent l'ensemble des 150 images pour lesquelles nous avons nous-mêmes collecté des scores, Ito *et al.* et Grühn et Scheibe n'ont en commun avec nous que 103 et 97 images, respectivement.

Les scores de valence que nous avons obtenus sont fortement corrélés à ceux des études précédentes ($.89 < r < .95$) : les participants à notre étude étaient fortement en accord avec les participants des études précédentes pour dire qu'une image véhiculait une émotion plus ou moins positive ou négative. Les scores d'arousal que nous avons obtenus sont modérément corrélés avec ceux des études précédentes ($.55 < r < .72$). Il faut noter que c'est également le cas pour les scores des précédentes études entre elles. Cela peut suggérer différentes hypothèses. Il est, par exemple, possible qu'une image suscite une plus grande variété de réactions dans la dimension d'arousal que de valence. Plus simplement, il est également possible que l'arousal soit un concept plus difficilement identifiable et estimable que la valence par les individus. D'autre part, cette corrélation modeste entre les scores d'arousal des différentes études pourrait s'expliquer par les différences (en matière de méthode, de population, d'âge des participants, etc.) entre les différentes études.

Pour déterminer si cette corrélation modeste entre les scores d'arousal existait également entre les différents participants à notre étude, nous avons utilisé une méthode employée par Isola *et al.* pour déterminer le degré de cohérence inter-individuelle de la mémorabilité des images (Isola et al., 2014). Suivant cette méthode, nous avons séparé aléatoirement nos participants en deux, et corrélé les scores d'arousal et de valence calculés sur la première moitié des participants avec ceux calculés sur la seconde moitié. Nous avons répété 25 fois cette opération (en divisant, donc, 25 fois nos participants en deux), et obtenu une corrélation moyenne de $r = .92$ pour l'arousal et de $r = .96$ pour la valence. Les résultats montrent que les participants de notre étude étaient fortement en accord s'agissant de l'arousal comme de la valence suscités par les 150 images utilisées. Cela suggère que la faible corrélation entre les scores d'arousal des différentes études est due à des facteurs (méthode employée, origine des participants, époque où les images ont été évaluées, etc.) qui diffèrent entre ces études — facteurs qui n'influencent pas, ou peu, l'évaluation de la valence.

	1	2	3	4	5	6	7	8	9
		Valence							
1. Notre étude									
2. Lang *et al.*, 1998	.95**								
3. Ito et *al.*, 1998	.93**	.96**							
4. Libkuman *et al.*, 2007	.89**	.90**	.87**						
5. Grühn et Scheibe, 2008	.95**	.92**	.92**	.89**					
		Arousal							
6. Notre étude	−.52**	−.55**	−.57**	−.49**	−.49**				
7. Lang *et al.*, 1998	−.32**	−.31**	−.34**	−.34**	−.23*	.69**			
8. Ito *et al.*, 1998	−.56**	−.52**	−.50**	−.48**	−.53**	.72**	.81**		
9. Libkuman *et al.*, 2007	−.17*	−.18*	−.22*	−.23**	−.18	.55**	.61**	.51**	
10. Grühn et Scheibe, 2008	−.86**	−.83**	−.84**	−.81**	−.93**	.69**	.50**	.67**	.36**

TABLE 6.1 – Matrice de corrélation des scores d'arousal et de valence de notre étude et des études précédentes pour les 150 images de l'IAPS utilisées. Les corrélations avec les scores de Ito *et al.* (1998) et de Grühn et Scheibe (2008) ne portent respectivement que sur les 103 et 97 images que nous avons en commun avec ces auteurs. $^{*} < .05$; $^{**}p < .01$.

6.2 Nature de la relation géométrique entre l'arousal et la valence

Plusieurs modèles théoriques supposent que l'arousal et la valence sont des dimensions indépendantes orthogonales (notamment (Barrett and Russell, 1999, Carver and Scheier, 1990, Larsen and Diener, 1992)). Quelques auteurs ont trouvé que l'arousal covariait positivement avec la valence (p. ex. (Pettinelli, 2008)) ou, au contraire, négativement (p. ex. (Tsai et al., 2006)). D'autres auteurs ont trouvé que l'arousal et la valence sont liées par une relation symétrique en forme de V ou de U (p. ex. (Jennings et al., 2000, Bernat et al., 2006, Bradley et al., 2001)) ou asymétrique, avec les stimuli négatifs associés à un arousal plus fort que les stimuli positifs (p. ex. (Ito and Cacioppo, 2005, Ito et al., 1998, Baumeister et al., 2001)). Le type de stimuli et la méthode de mesure sont donc susceptibles d'influencer la forme de la relation géométrique observée entre l'arousal et la valence.

Les patterns constatés par les auteurs ayant fait évaluer des images de l'IAPS montrent généralement une relation en forme de U ou de V entre l'arousal et la valence, lorsque les images sont projetées dans un espace cartésien suivant leurs scores de valence (portés en abscisse) et d'arousal (portés en ordonnée) (Backs et al., 2005, Ito et al., 1998, Lang et al., 1997, Lang et al., 2008, Libkuman et al., 2007). Dans ces études, les images très

positives et très négatives étaient typiquement évaluées comme hautement activatrices ; et, plus la valence évaluée tendait vers la neutralité (i.e. lorsque la note de valence de l'image se rapprochait de 5, soit le milieu des échelles de valence utilisées dans ces études), plus la note d'arousal attribuée aux images tendait à être faible. Certains auteurs n'ont toutefois pas obtenu un tel pattern (Grühn and Scheibe, 2008, Ribeiro et al., 2005). Dans ces études, les images de l'IAPS positives n'ont pas été évaluées comme plus activatrices que les images neutres.

Dans notre étude, nous avons trouvé une relation asymétrique entre les scores d'arousal et de valence, présentée dans la figure 6.1 : les images négatives ont typiquement été évaluées par nos participants comme suscitant un arousal élevé, les images neutres un arousal faible et les images positives un arousal modéré. Nous avons observé une corrélation linéaire négative entre l'arousal et la valence ($r = -.52, p < .001$). Après inspection du nuage de points, nous avons décidé de calculer séparément la corrélation entre l'arousal et la valence positive, et la corrélation entre l'arousal et la valence négative. Dans cet objectif, nous avons séparé les images dont les scores de valence étaient inférieurs à la valeur 5 (qui représente la neutralité sur l'échelle SAM de valence) de celles dont les scores étaient supérieurs à 5. Nous avons observé une association linéaire positive entre l'arousal et la valence positive ($r = .77, p < .0001$), et une association linéaire négative entre l'arousal et la valence négative ($r = -.92, p < .0001$). Ces résultats suggèrent que la valence négative est plus fortement associée à l'arousal que la valence positive.

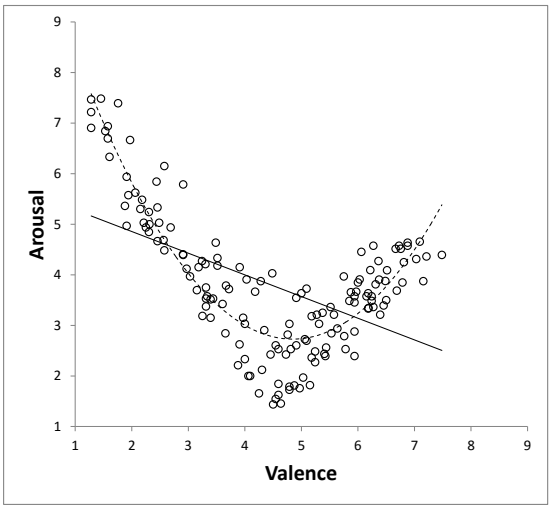


FIGURE 6.1 – Nuage de points représentant les 150 images de l'IAPS notées dans notre étude en fonction de leurs scores de valence et d'arousal. Les associations linéaires et quadratiques entre les scores d'arousal et de valence sont représentées par des lignes continue et pointillée, respectivement.

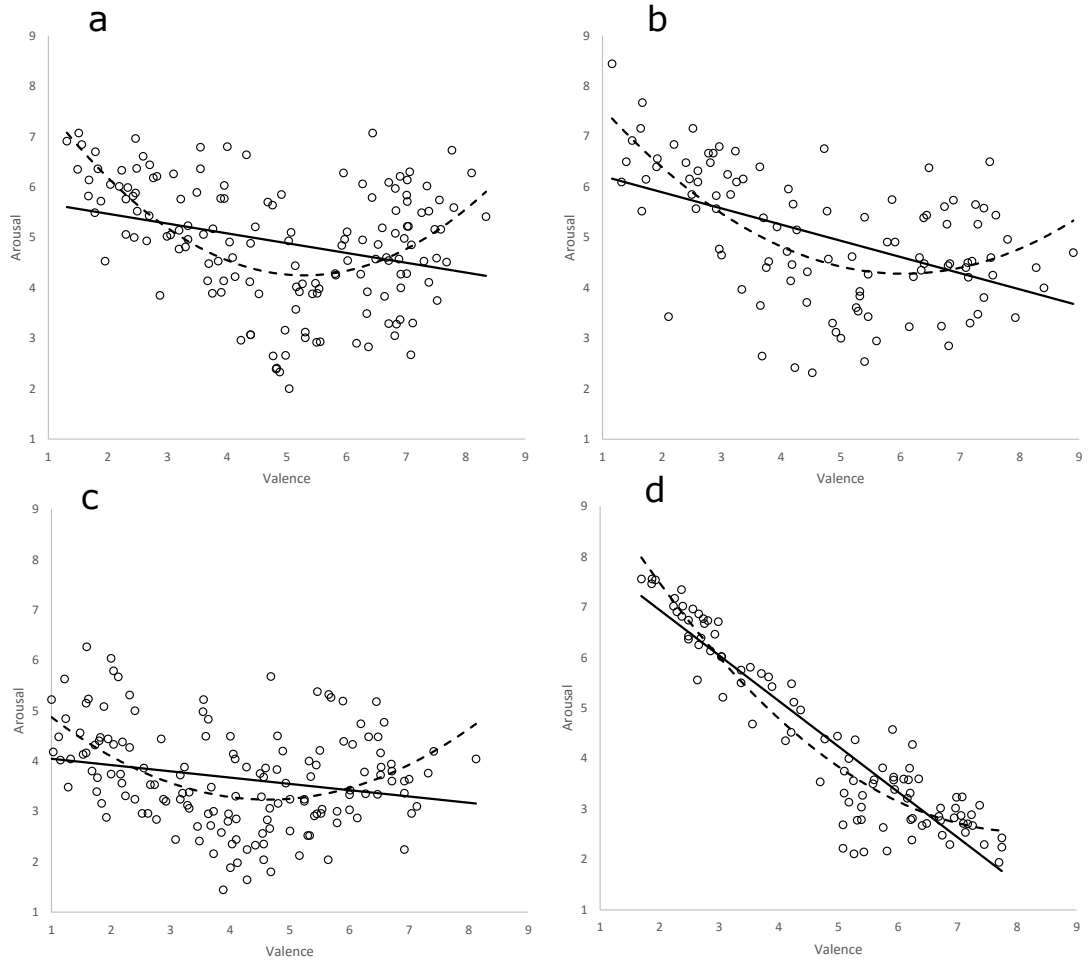


FIGURE 6.2 – Nuages de points des scores d'arousal et de valence pour les images de l'IAPS utilisées dans notre études qui ont été notées lors d'études précédentes : (a) Lang *et al.*, 1998 (150 images en commun) ; (b) Ito *et al.*, 1998 (103 images) ; Libkuman *et al.* (150 images) ; Grühn et Scheibe (97 images). Les associations linéaires et quadratiques entre les scores d'arousal et de valence sont représentées par des lignes continues et pointillées, respectivement.

6.3 Valence et arousal moyens

La table 6.2 présente les scores moyens d'arousal et de valence (avec les écart-types) obtenus dans notre étude et dans les études précédentes (Grühn and Scheibe, 2008, Ito et al., 1998, Lang et al., 1997, Libkuman et al., 2007), pour les 65 images de l'IAPS que l'ensemble de ces auteurs ont fait évaluer. Pour le sous-ensemble d'images comparées, nous pouvons voir dans la table 6.2 que nos scores de valence comme d'arousal sont de valeur intermédiaire lorsqu'on les compare aux scores des études précédentes.

	Valence		Arousal	
Source de la notation	*M*	*SD*	*M*	*SD*
Notre étude	4.40	1.77	3.97	1.34
Lang *et al.*, 1997	4.97	2.12	5.05	1.08
Ito *et al.*, 1998	5.00	2.14	4.99	1.22
Libkuman *et al.*, 2007	4.05	1.89	3.63	0.98
Grühn et Scheibe, 2008	4.79	1.90	4.49	1.81

TABLE 6.2 – Statistiques descriptives pour 65 images de l'IAPS

Un test de Kruskal-Wallis et un test HSD de Tukey confirment en partie cette observation. Pour l'arousal, l'effet principal est significatif ($X_2(4) = 52.61, p < .000$). Le test de Tukey révèle que nos scores diffèrent significativement des scores obtenus par (Lang et al., 1997) et (Ito et al., 1998), mais pas des scores obtenus par (Libkuman et al., 2007) et (Grühn and Scheibe, 2008). Pour la valence, l'effet principal est également significatif ($X_2(4) = 12.80, p < .01$). Le test de Tukey ne révèle cependant aucune différence significative entre les scores obtenus dans notre étude et ceux des études précédentes considérées. L'effet principal du test de Kruskal-Wallis est dû à la différence significative entre les scores de Libkuman *et al.* et les scores des autres études précédentes. Bien sûr, la non-significativité d'un effet ne peut être interprété comme la non-existence d'un tel effet. Ce que nous pouvons dire de ces résultats, essentiellement, est que nos participants ont eu tendance à noter l'arousal suscité par les images plus faiblement que les participants de Lang *et al.* et de Ito *et al.*.

6.4 Discussion

Les scores d'émotion que nous avons obtenus dans notre étude ont été comparés avec les scores rapportés pour les mêmes images par plusieurs études (Lang et al., 1997, Ito et al., 1998, Libkuman et al., 2007, Grühn and Scheibe, 2008). Globalement, les évaluations réalisées dans les différentes études corrèlent assez fortement, en particulier pour la valence. Nous avons cependant trouvé des différences, qui méritent d'être discutées.

Les analyses statistiques des scores d'arousal et de valence montrent que nos participants ont globalement évalué les images comme suscitant moins d'arousal que les participants de (Lang et al., 1997) et (Ito et al., 1998). Une raison pourrait en être que nos participants évaluaient des images qu'ils avaient déjà vues deux fois précédemment, dans les tâches de reconnaissance, alors que dans ces deux études les participants évaluaient des images qu'ils voyaient pour la première fois. Le fait de revivre une expérience avec le même matériel émotionnel pourrait avoir diminué l'intensité de l'arousal qu'il a suscité. Les résultats obtenus dans (Libkuman et al., 2007) montrent également que les participants à cette étude ont globalement attribué des notes d'arousal plus faibles que dans ceux de ces deux précédentes études. Dans cette étude, plusieurs images étaient vues et évaluées deux fois par les mêmes participants sur des dimensions relatives à l'arousal. Les auteurs ont observé que la moyenne d'arousal était significativement plus basse pour la seconde évaluation que pour la première. Selon eux, leur résultat pourrait s'expliquer par le fait que le nombre d'images évaluées pourrait avoir eu une influence sur l'évaluation de l'arousal, dans le sens où les images auraient progressivement été jugées moins activatrices au fur et à mesure que le nombre d'évaluations augmentait. Dans notre étude, les images n'étaient évaluées qu'une seule fois ; cela suggère que ce pourrait être, non pas seulement la ré-évaluation des images, mais simplement le fait de les revoir, qui explique la baisse de l'intensité de l'arousal observée par Libkuman *et al.*.

Si la présentation répétée d'une même image tend à diminuer l'intensité de l'arousal perçu par un observateur, il est possible que la présentation d'images suscitant des émotions similaires tende également à diminuer l'intensité de l'arousal perçu par un observateur lors du visionnage de ces images. À ce propos, nous ne pouvons exclure la possibilité que la probable multiplication de la consommation d'images violente dans la société occidentale actuelle n'ait pas modifié notre sensibilité à ce type d'images. Les scores d'arousal proposés par (Lang et al., 1997) et (Ito et al., 1998), desquels nos scores d'arousal diffèrent significativement, ont été obtenus dans des années 1990. Le fait que les scores d'arousal obtenus par (Libkuman et al., 2007) et (Grühn and Scheibe, 2008) soient également significativement plus bas que ceux obtenus dans ces deux études est cohérent avec cette hypothèse. En somme, une tendance semble se dégager : les évaluations les plus récentes de l'arousal suscité par les images de l'IAPS tendent globalement à leur attribuer des scores plus bas. Cependant, les différences de méthode entre ces différentes études doivent éveiller notre méfiance quant à cette explication ; de nouveaux travaux sont nécessaires pour confirmer cette hypothèse.

Alors que nous avons trouvé une différence significative entre les scores d'arousal de notre étude et ceux obtenus dans (Lang et al., 1997) et (Ito et al., 1998), nous n'avons pas observé de telle différence pour les scores de valence, avec aucune des études auxquelles nous nous sommes intéressés (Lang et al., 1997, Ito et al., 1998, Libkuman et al., 2007, Grühn and Scheibe, 2008). Cela ne signifie, bien sûr, pas qu'une telle différence n'existe

pas pour la valence, mais suggère que l'arousal pourrait être davantage influencé que la valence par les facteurs qui diffèrent entre notre étude et celles de Lang *et al.* et de Ito *et al.*. En particulier, pour reprendre l'hypothèse évoquée précédemment, l'arousal pourrait être plus affecté que la valence par le fait que les images soient vues plusieurs fois.

Nous avons trouvé des corrélations modérément importante entre les scores d'arousal obtenus dans notre étude et ceux obtenus dans les études précédentes, et des corrélations fortes entre les scores de valence obtenus dans notre étude et ceux obtenus dans les études précédentes. C'est également le cas pour les études précédentes entre elles, comme nous pouvons le voir dans la table 6.1 : les corrélations sont, dans tous les cas, plus fortes pour les scores de valence que pour les scores d'arousal. Au premier regard, cela pourrait suggérer qu'une image suscite chez les individus une plus grande variété de réactions dans la dimension d'arousal que dans la dimension de valence, puisque les participants des différentes études sont fortement en accord sur la positivité/négativité des images, et moins sur le degré d'arousal qu'elles suscitent. Cependant, nous avons trouvé que l'accord inter-observateur était très fort dans notre étude, aussi bien pour la valence que pour l'arousal, ce qui suggère que les différences entre les études (méthodes variées, participants de différentes nationalités, époques différentes, etc.) expliquent les différences observées dans les scores d'arousal.

6.5 Conclusion

Il ressort de ce chapitre trois points principaux. Premièrement, les scores d'émotion obtenus dans l'étude présentée dans le chapitre 5 sont cohérents. Ensuite, le degré de généralisabilité des scores d'arousal et de valence pour des images est assez important, en particulier pour la valence. Ce point est important pour la prédiction computationnelle de l'émotion véhiculée par les images, qui porte généralement sur l'information émotionnelle intrinsèque des images, sans considération des facteurs contextuels ou individuels susceptibles d'influencer les émotions suscitées par les images (Joshi et al., 2011). Indirectement, il est également important pour la prédiction computationnelle de la mémorabilité des images, puisque, comme on l'a précédemment évoqué, l'information émotionnelle algorithmiquement extractible des images pourrait être mise à profit d'un tel objectif. Enfin, nos résultats confirment que la relation entre l'arousal et la valence n'est pas géométrique. En particulier, notre étude renforce la position selon laquelle les images négatives tendent à susciter plus d'arousal que les images positives.

Dans le chapitre suivant, nous étudions conjointement les scores de mémorabilité et d'émotion obtenus dans notre base de données, et nous intéressons à l'évolution de la mémorabilité des images en mémoire à long terme.

7

Influence de l'émotion sur la mémorabilité des images

Les images qui suscitent une réaction émotionnelle significative sont généralement mieux retenues que les images émotionnellement neutres ; nous l'avons établi dans le chapitre 3. D'autre part, une image mémorable quelques minutes après son encodage n'est peut-être plus mémorable le lendemain. Le processus de consolidation des souvenirs, influencé par un certain nombre de facteurs, au premier rang desquels l'émotion suscitée par l'image et le sommeil, aura peut-être gravé dans la nuit une image dans notre mémoire à long terme, ou peut-être l'aurons nous oubliée demain. À partir des scores de mémorabilité et d'émotion que nous avons obtenus dans l'expérience présentée dans le chapitre 5, nous étudions dans ce chapitre la mémorabilité des images, et l'influence de la durée de rétention et de l'émotion sur celle-ci.

Nous comparons également nos scores de mémorabilité avec ceux rapportés précédemment par (Grühn and Scheibe, 2008) et (Libkuman et al., 2007) pour certaines des images de l'IAPS que nous avons utilisées. Comparer nos score de mémorabilité avec ceux de (Grühn and Scheibe, 2008), également obtenus à l'aide d'une tâche de reconnaissance, mais dans des conditions expérimentales différentes des nôtres, nous donnera un aperçu de la mesure dans laquelle nos scores sont généralisables. Quant à la comparaison de nos scores de mémorabilité à ceux de (Libkuman et al., 2007), elle nous permettra d'évaluer la capacité humaine à prédire la mémorabilité des images, qui détermine, en fait, si l'annotation manuelle pour obtenir des scores de mémorabilité pour des images est envisageable (nous avons expliqué en détail cette problématique dans la section 4.2.3 du chapitre 4).

Dans ce chapitre, nous rapportons d'abord les résultats relatifs à ces différents points. Puis, en fin de chapitre, nous discutons ces résultats.

7.1 Comparaison de nos scores de mémorabilité avec les études antérieures

Dans l'expérience décrite au chapitre 5, nous avons obtenu des scores de mémorabilité pour 150 images de l'IAPS. La mémorabilité d'images de l'IAPS a, cependant, déjà été le sujet, secondaire, de deux études antérieures à la nôtre (Grühn and Scheibe, 2008,Libkuman et al., 2007). Les auteurs de ces études ont rapporté des scores de mémorabilité pour un certain nombre d'images de l'IAPS, dont certaines font partie des 150 images de notre base de données. La *mémorabilité* recouvre cependant des réalités différentes dans nos études respectives. Dans notre étude et celle de (Grühn and Scheibe, 2008), les scores de mémorabilité ont été calculés à partir des résultats à une tâche de reconnaissance. Nos protocoles expérimentaux sont cependant assez différents. Dans (Libkuman et al., 2007), la mémorabilité correspond à un jugement porté a priori sur la mémorabilité des images, et non à une mesure objective d'une performance de mémoire, telle qu'on peut en obtenir en utilisant une tâche de reconnaissance. Comparer nos scores de mémorabilité à ceux de ces auteurs nous permet d'évaluer notre capacité à prédire la mémorabilité des images, comme nous l'avons expliqué en détail dans la section 4.2.3 du chapitre 4.

La table 7.1 présente les corrélations entre nos scores de mémorabilité et ceux rapportés par ces auteurs, pour les images que nous avons en commun.

7.1.1 Généralisabilité de nos scores de mémorabilité

Les scores de mémorabilité de Grühn et Scheibe correspondent, comme on l'a dit, à une performance de reconnaissance : les participants devaient indiquer si une image affichée avait été vue dans une tâche d'encodage effectuée quelques minutes auparavant (pour les détails de la tâche de reconnaissance, voir (Grühn et al., 2007)). Dans la table 7.1, sont présentées séparément les corrélations de nos scores de mémorabilité avec ceux de Grühn et Scheibe pour les jeunes adultes, et pour l'ensemble de leurs participants (adultes jeunes et âgés). En effet, ces auteurs ont montré des différences de mémorabilité entre ces deux groupes de participants. Nos participants ($18-41$ ans ; $\bar{x} = 22.54, SD = 5.01$) étaient approximativement de l'âge de leurs jeunes adultes ($18 - 31$ ans ; $\bar{x} = 25.23, SD = 3.39$).

La table 7.1 montre que nos scores de mémorabilité correspondant au premier test de mémoire[1] (effectué, à l'instar de (Grühn and Scheibe, 2008), quelques minutes après

[1]Dans la suite de ce chapitre, nous utiliserons la dénomination « T1 » pour qualifier les scores de

TABLE 7.1 – Corrélations entre les scores de mémorabilité obtenus dans notre étude et les scores de mémorabilité des études précédentes.

	1	2	3	4
1. Cohendet *et al.*, Jour 1				
2. Cohendet *et al.*, Jour 2	.59**			
3. Grühn et Scheibe, 2008 (jeunes adultes)	.39**	.36**		
4. Grühn et Scheibe, 2008 (tous les participants)	.41**	.33**	.84**	
5. Libkuman *et al.*, 2007	.17*	.25**	.09	.00

Note – Les corrélations avec les scores de Libkuman *et al.* (2007) et de Grühn et Scheibe (2008) ont été calculées sur les 150 images et 97 images que nous avions en commun avec ces auteurs, respectivement. $^{*}p < .05$. $^{**}p < .01$.

l'encodage) sont modérément corrélés avec les scores de mémorabilité de Grühn et Scheibe pour les jeunes adultes ($r = .39$) et pour l'ensemble de leurs participants ($r = .41$).

Les modalités de mesure dans notre étude différaient de celles de Grühn et Scheibe (p. ex. durée de présentation des images, difficulté de la tâche, etc.), ce qui pourrait expliquer que la corrélation observée entre les scores de mémorabilité obtenus dans nos deux études est modérée. Pour tester cette hypothèse, nous avons utilisé la méthode proposée par Isola *et al.* pour déterminer le degré de cohérence inter-individuelle de la mémorabilité des images (Isola et al., 2014). Suivant cette méthode, nous avons divisé 25 fois nos participants en deux groupes aléatoires de taille équivalente, et avons calculé à chaque fois la corrélation entre les scores de mémorabilité obtenue par une moitié des participants avec ceux obtenus par l'autre moitié. En suivant ce procédé, nous avons obtenu une corrélation ρ de Spearman moyenne de $.70$ (et une corrélation r de Pearson moyenne de $.66$). Ce résultat est proche de celui rapporté par (Isola et al., 2014)), qui ont trouvé une corrélation moyenne de Spearman de $.75$ en employant la même méthode de calcul. Il confirme que les images les plus mémorables pour un groupe d'individus sont souvent les images les plus mémorables pour un autre groupe d'individus.

mémorabilité qui, dans l'expérience présentée dans le chapitre 5, ont été obtenus à partir des résultats au test de mémoire effectué quelques minutes après l'encodage des images. La dénomination « T2 », quant à elle, sera utilisée pour qualifier les scores de mémorabilité obtenus à partir des résultats du test de mémoire effectué un jour après l'encodage des images.

7.1.2 Capacité humaine à prédire la mémorabilité d'une image : le rôle de l'arousal

Ce que Libkuman *et al.* (2007) appellent mémorabilité d'une image ne coïncide pas avec l'acception qui lui est donnée dans notre base de données. En effet, leurs scores de mémorabilité ne correspondent pas à une mesure de performance mnésique, mais à des jugements subjectifs portés a priori sur le degré de mémorabilité d'images. Les corrélations entre nos scores de mémorabilité et les leurs sont présentées dans la table 7.1. Nos scores de mémorabilité T1 sont faiblement corrélés avec ceux obtenus par (Libkuman et al., 2007) ($r = .17$), tout comme nos scores de mémorabilité T2 ($r = .25, p < .01$). Ces corrélations, quoique faibles, suggèrent que les personnes sont capables de prédire, dans une certaine mesure, la mémorabilité réelle (i.e. mesurée objectivement) d'une image lorsqu'on leur présente cette image pour la première fois. Ce résultat est contraire (Isola et al., 2014), mais en accord avec (Underwood, 1966, Jacob and Nelson, 1990).

Libkuman *et al.* (2007) ont trouvé une corrélation de $.65$ entre les scores de mémorabilité et d'arousal obtenus dans leur étude. Dans notre étude, l'arousal est corrélé à la performance de mémoire (nous le montrerons dans la section 7.2). Nous nous sommes demandés si l'arousal suscité par une image était un indice fiable pour juger du degré de mémorabilité d'une image. Pour répondre à cette question. Pour répondre à cette question, nous avons d'abord transformé les scores de Libkuman *et al.*, qui correspondent à des moyennes de valeurs sur une échelle en neuf points, en pourcentages, pour qu'ils soient comparables à nos scores de mémorabilité. Ensuite, nous avons calculé pour chacune des 150 images de notre base de données un Δ_{T1}, en soustrayant au score de mémorabilité T1 le score de mémorabilité obtenu pour cette même image par Libkuman *et al.*. Nous avons répété la même opération pour les scores de mémorabilité calculés à partir des résultats obtenus au second test de mémoire (réalisé un jour après l'encodage des images), et obtenu pour chaque image un Δ_{T2}.

Nous avons observé une corrélation négative modérée entre les scores d'arousal et les Δ_{T1} ($r = -.45, p < .001$), et une corrélation négative faible entre les scores d'arousal et les Δ_{T2} ($r = -.22, p < .01$). Ces résultats signifient que plus l'arousal moyen véhiculé par une image était fort, moins la différence entre la mémorabilité réelle (i.e. mesurée objectivement dans notre étude) et la mémorabilité prédite (i.e. telle qu'elle a été jugée par les participants dans (Libkuman et al., 2007)) était importante. Il faut préciser ici que l'ensemble des Δ étaient positifs, c'est-à-dire que la mémorabilité réelle était, pour toutes les images, plus élevée que la mémorabilité prédite. En somme, nos résultats suggèrent que l'arousal suscité par une image améliore la précision du jugement porté a priori sur sa mémorabilité.

7.2 Arousal, valence et mémorabilité des images

L'émotion qu'une image véhicule joue un rôle essentiel dans la mémorisation des images et leur préservation en mémoire à long terme, comme nous l'avons établi dans le chapitre 3. En particulier, les images émotionnellement chargées sont généralement mieux retenues que les images émotionnellement neutres (Bradley et al., 1992). Pour vérifier si cette vérité s'appliquait aux images de notre base de données, nous avons analysé conjointement les scores de mémorabilité, d'arousal et de valence, obtenus grâce à l'expérience présentée dans le chapitre 5.

Nous avons observé une corrélation linéaire positive entre les scores d'arousal et les scores de mémorabilité de nos 150 images, à la fois pour la mémorabilité T1 ($r = .23, p < .01$), et pour la mémorabilité T2 ($r = .42, p < .001$). Nous avons également observé une corrélation linéaire négative entre les scores de valence et les scores de mémorabilité T1 ($r = -.27, p < .001$) et T2 ($r = -.28, p < .001$). Ces résultats montrent que l'arousal et la valence sont liés à la reconnaissance correcte d'images préalablement étudiées, que celle-ci soit mesurée quelques minutes après l'encodage du matériel ou un jour après.

Un point actuellement peu clair concerne la probabilité relative des images négatives et positives d'être récupérés en mémoire, comme nous l'avons expliqué dans la section 3.2.1 du chapitre 3. Pour investiguer cette question, après l'analyse des nuages de points présentés dans la figure 7.1, nous avons décidé de tester l'association entre la mémorabilité et les scores de valence pour les images de valence positive, d'une part, et pour les images de valence négative, d'autre part. Dans cet objectif, nous avons soustrait à chacun des scores de valence la valeur correspondant à la neutralité sur l'échelle SAM de la valence (i.e. 5), ce qui nous a permis d'obtenir deux groupes d'images : d'un côté, les images négatives, et de l'autre les images positives. Pour les images négatives, nous avons observé une association linéaire négative entre les scores de valence et de mémorabilité T1 ($r = -.20, p < .05$) et T2 ($r = -.42$, $p < .001$). En revanche, pour les images positives, nous n'avons pas observé de corrélation linéaire significative entre les scores de valence et de mémorabilité T1 ($r = -.15, p = .22$) ou T2 ($r = .14, p = .27$). Pour tester l'hypothèse selon laquelle la mémorabilité moyenne des images négatives différait de la mémorabilité moyenne des images positives, nous avons utilisé un test de Student. Pour la mémorabilité T1, les moyennes des groupes « images de valence négative » et « images de valence positive » étaient 0.80 et 0.71, respectivement ; la moyenne des deux groupes différait significativement ($t = 3.54, n_1 = 87, n_2 = 63, p < .001$, *bilatéral*). Pour la mémorabilité T2, les moyennes des groupes « images de valence négative » et « images de valence positive » étaient 0.58 et 0.51, respectivement ; la moyenne des deux groupes différait également significativement ($t = 2.00 n_1 = 87, n_2 = 63, p < .05$, *bilatéral*). Ces résultats signifient que les images négatives étaient mieux reconnues que les images positives, dans les deux tests de mémoire proposés à nos participants dans l'expérience décrite dans le

chapitre 5.

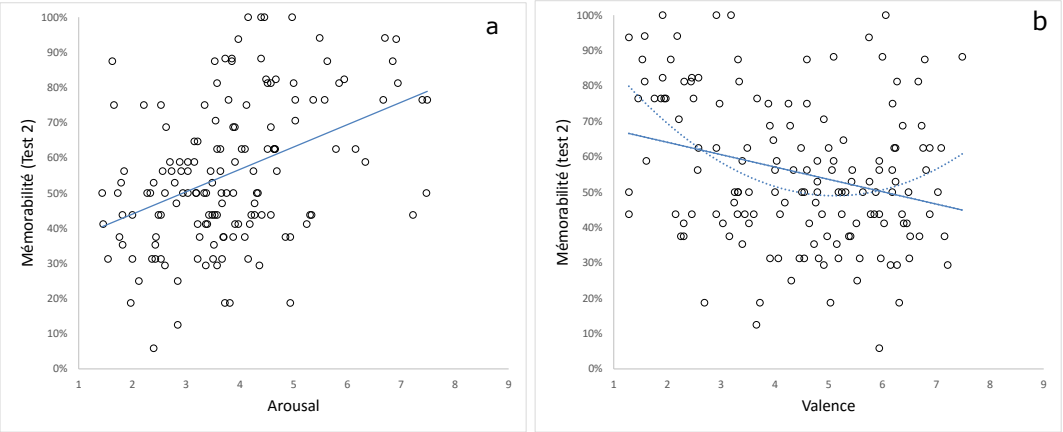


FIGURE 7.1 – Nuages de points des 150 images notées dans notre étude en fonction de leurs scores de mémorabilité calculés à partir des résultats au second test de mémoire (effectué un jour après l'encodage) et (a) de leurs scores d'arousal et (b) de leurs scores de valence. Les associations linéaires et quadratique entre les scores d'émotion et de mémorabilité sont représentées par des lignes solides et pointillée, respectivement.

7.3 Persistance de la mémorabilité des images

En accord avec la théorie de la consolidation, le stockage de l'information en mémoire ne serait pas définitif à la fin de la phase d'encodage : il faudrait un certain au cerveau pour stabiliser la trace de mémoire (McGaugh, 2000). Durant cette phase de consolidation, les souvenirs seraient fragiles et pourraient être facilement altérés ou modifiés : la trace de mémoire d'un évènement pourrait être renforcée, ou au contraire affaiblie (Brosch et al., 2013). C'est un point important puisqu'il s'ensuit, comme nous l'avons expliqué dans la section 1.2.2 du chapitre 1, qu'une image mémorable quelques minutes après son encodage pourrait ne plus être mémorable après un certain temps.

D'autre part, la consolidation pourrait ne pas s'appliquer uniformément à l'ensemble des images. En effet, comme nous l'avons évoqué dans la section 3.1.2 du chapitre 3, l'émotion — et en particulier l'arousal — agirait sur le processus consolidation en conduisant à un renforcement de la trace de mémoire (LaBar and Phelps, 1998, Dolcos et al., 2004). Par conséquent, les images émotionnellement chargées pourraient être mieux préservées que les images émotionnellement neutres pendant la phase de stockage mnésique. Il s'ensuivrait que l'ordre de mémorabilité des images d'une base de données — par exemple, de la base de (Isola et al., 2011b) — pourrait être bouleversé par l'augmentation de la durée de rétention des images, en raison des émotions différentes véhiculées par ces images.

Dans l'expérience décrite dans le chapitre 5, nous avons mesuré la mémoire de nos participants à deux moments différents : après une durée de rétention mnésique de quelques minutes, puis un jour après. À partir des résultats obtenus à ces deux tests de mémoire, nous avons calculé deux scores de mémorabilité, T1 et T2, pour chaque image de notre base de données, qui nous permettent d'investiguer ces questions.

7.3.1 Les images les plus mémorables après quelques minutes ne sont pas les images les plus mémorables un jour après

Pour déterminer si les scores de mémorabilité T1 et T2 de nos 150 images différaient significativement, nous avons utilisé un test de Student. La moyenne des scores de mémorabilité des 150 images pour T1 et T2 était de 0.76 et 0.55, respectivement ; la moyenne des deux groupes différait significativement ($t = 9.51, p < .0001$). Ce résultat signifie que la mémorabilité des images quelques minutes après leur encodage était significativement plus haute que la mémorabilité de ces mêmes images un jour après leur encodage. Cela suggère que les scores de mémorabilité de (Isola et al., 2011b), calculés à partir des résultats à un test de mémoire effectué quelques minutes après l'encodage des images, ne reflètent pas, ou seulement en partie, la mémorabilité des images de leur base de données après une durée de rétention plus longue.

Pour tester l'hypothèse selon laquelle les images les plus mémorables quelques minutes après l'encodage ne sont pas nécessairement les images les plus mémorables un jour après, nous avons calculé un coefficient de corrélation de Spearman entre les scores de mémorabilité T1 et les scores de mémorabilité T2. La corrélation observée, positive, est modérée : $\rho = .58, p < .0001$. Ce résultat montre qu'après l'écoulement d'une durée de rétention d'un jour en mémoire à long terme, l'ordre de mémorabilité des images a changé. Ce résultat suggère que le fait qu'une image soit mémorable quelques minutes après l'encodage ne garantit pas qu'elle soit mémorable après un délai de rétantion plus important. Les implications de ce résultat sont importantes puisque, à notre connaissance, l'ensemble des études ayant porté sur la mémorabilité des images en vision par ordinateur ont utilisé des scores de mémorabilité correspondant à des performances de mémoire mesurées quelques minutes après l'encodage des images. Or, l'intérêt de prédire une mémorabilité qui n'est valable que pendant une durée relativement courte n'a que peu d'intérêt, en particulier eu égard aux applications envisagées par (Isola et al., 2011b, Isola et al., 2014) pour les modèles prédictifs (p. ex. la création de matériel éducatif, le design d'interfaces utilisateur, l'aide aux personnes âgées, etc.).

7.3.2 Influence de l'émotion sur la baisse de mémorabilité des images

L'analyse des résultats présentée dans la section 7.2 a montré la corrélation entre les scores d'arousal et de mémorabilité de nos 150 images est plus forte à T1 qu'à T2. En

revanche, la corrélation entre les scores de valence et de mémorabilité demeure la même à T1 et T2. Ces résultats corroborent l'hypothèse d'un rôle essentiel de l'arousal — et moindre de la valence — dans la consolidation des souvenirs d'images.

Pour tester la relation de l'arousal et la valence avec la baisse de mémorabilité advenue durant le délai de rétention mnésique (d'une journée) séparant le premier test de mémoire du second, nous avons calculé pour chacune de nos 150 images *i* un Δ tel que : $\Delta_{M_i} = M_{1_i} - M_{2_i}$, avec M_1 le score de mémorabilité de l'image calculé à partir des résultats au premier test de mémoire et M_2 le score de mémorabilité calculé à partir des résultats au second test (voir la figure 7.2). Ensuite, nous avons calculé la corrélation entre les scores d'arousal et les Δ_M pour les 150 images notées, et répété la même opération avec les scores de valence (voir la figure 7.2). Nous avons trouvé une corrélation linéaire négative entre les scores d'arousal et les Δ_M ($r = -.26, p < .005$), mais pas de corrélation significative entre les scores de valence et les Δ_M ($r = .06, p = .40$). Ces résultats confirment l'hypothèse d'un rôle important de l'arousal dans la préservation des souvenirs.

Après inspection visuelle des nuages de points (voir la figure 7.2), nous avons également décidé de tester séparément l'association entre les Δ_M et les scores de valence, pour les images positives et négatives. Nous avons observé une association linéaire positive entre les Δ_M et les scores de valence pour les images de valence négative ($r = .34, p < .01$), mais pas d'association significative pour les images de valence positive ($r = . - 00, p < .988$). Ces résultats suggèrent que plus les images sont négatives, plus elles tendent à être préservées de l'oubli en mémoire à long terme. Cependant, les dimensions d'arousal et de valence ne sont pas indépendantes, comme nous l'avons montré dans la section 6.2 du chapitre 6 : les images négatives de notre base de données tendent à susciter plus d'arousal que les images neutres et positives. Par conséquent, l'effet confondu de l'arousal pourrait expliquer notre résultat. C'est ce que suggère l'association quadratique entre la valence et l'écart de mémorabilité entre les deux tests de mémoire, tracée dans la figure 7.2, qui rappelle l'association quadratique entre les scores d'arousal et de valence (voir la figure 6.1).

7.4 Discussion

Dans les sections précédentes de ce chapitre, nous avons présenté des résultats en vue de répondre à différentes questions. D'abord, nous avons comparé les scores de mémorabilité que nous avons obtenus dans l'expérience décrite dans le chapitre 5 avec ceux d'études antérieures (Libkuman et al., 2007, Grühn and Scheibe, 2008), ce qui nous a permis d'estimer le degré de généralisabilité de nos scores et la capacité humaine à juger a priori de la mémorabilité des images. Ensuite, nous avons montré que les scores d'arousal et de valence de nos 150 images étaient liés à leurs scores de mémorabilité. Finalement, en comparant les scores de mémorabilité T1 et T2, nous avons quantifié la

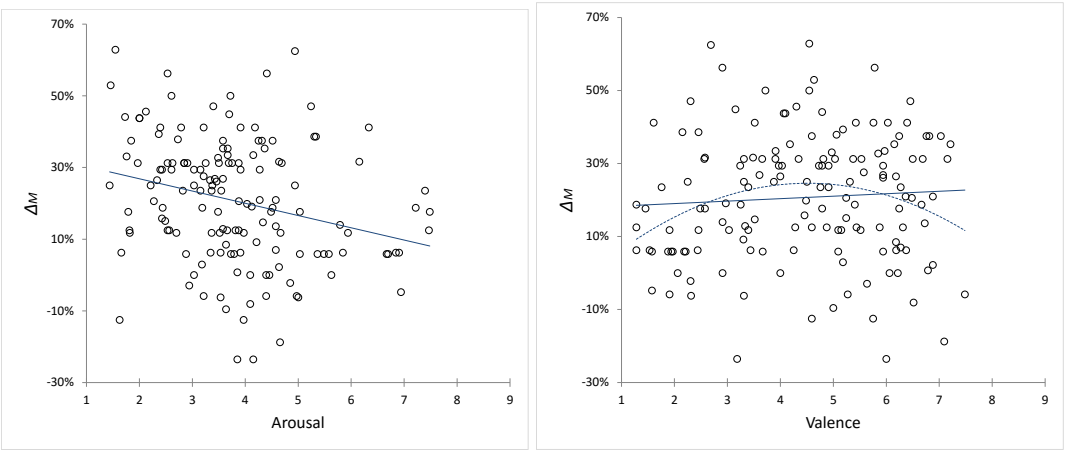


FIGURE 7.2 – Nuages de points des 150 images évaluées dans notre étude. Les images sont projetées dans l'espace cartésien selon leurs scores d'arousal (à gauche) ou de valence (à droite) et la baisse de leur mémorabilité durant la période de rétention mnésique d'un jour qui séparait les deux tests de mémoire. Les associations linéaires et quadratique entre la dimension émotionnelle considérée et l'écart de mémorabilité entre les deux tests de mémoire sont représentées par des lignes continues et pointillée, respectivement.

baisse de mémorabilité durant la période de rétention d'un jour qui séparait les deux tests de mémoire dans notre expérience, et avons montré que l'émotion véhiculée par les images était liée à celle-ci. Dans cette section, nous discutons ces différents résultats.

Les scores de mémorabilité dans notre étude et les précédentes

Les scores de mémorabilité des images notées dans notre étude ont été comparés avec ceux rapportés par deux études précédentes, pour les images que nous avions en commun (Grühn and Scheibe, 2008, Libkuman et al., 2007).

Nous avons trouvé une corrélation linéaire positive modérée entre nos scores de mémorabilité et ceux de Grühn et Scheibe. Nous avons également trouvé une forte concordance inter-observateur pour la mémorabilité dans notre étude, qui signifie qu'une image mémorable pour un participant était probablement mémorable pour un autre. Isola *et al.* sont également arrivé à la même conclusion (Isola et al., 2014). Ces deux résultats, pris ensemble, signifie que la corrélation modérée observée entre nos scores de mémorabilité et ceux de Grühn et Scheibe est due à des différences entre nos deux études. Ces différences portent probablement sur nos protocole expérimentaux, puisque l'étude de Grühn et Scheibe est assez récente et que leurs participants « jeunes adultes » sont de culture européenne et d'âge proche de celui de nos participants, ce qui permet d'évacuer les facteurs âge, culture et époque, susceptibles d'influencer la mémorabilité des images.

En particulier, la manière de mesurer la mémoire différait dans nos deux études. Dans Grühn et Scheibe, les images à reconnaître étaient présentées pour la moitié d'entre elles dans des conditions d'émotion homogènes, tandis que dans notre étude des images véhiculant une large palette d'émotions étaient présentées lors des tâches de reconnaissance. Le contexte de récupération des images a pu avoir un impact sur les performances de mémoire mesurées. La durée de présentation des images variait également entre nos deux études, ainsi que la difficulté de la tâche ; on ne peut exclure que ces facteurs aient eu un effet sur la mémorabilité relative des images entre elles.

Dans Libkuman *et al.*, il était demandé aux participants de juger a priori du degré de mémorabilité d'images tirées de l'IAPS, en s'aidant d'une échelle d'auto-évaluation. Les scores de mémorabilité obtenus par Libkuman *et al.* ne correspondent donc pas à une performance moyenne de mémoire, mais peut être considérée comme une mesure de méta-mémoire : un jugement fondé sur ce dont on pense qu'on se souviendra. Le jugement fondé sur l'impression qu'on va se souvenir (c'est-à-dire la mémorabilité) peut être rapproché du jugement porté sur la facilité d'apprentissage du matériel, comme nous l'avons expliqué dans la section 4.2.3 du chapitre 4. Il a été montré que les jugements portés sur la facilité d'apprentissage d'un matériel prédisaient la performance de rappel subséquente (Underwood, 1966, Leonesio and Nelson, 1990). La recherche sur la méta-mémoire a cependant surtout porté sur du matériel verbal. Une étude plus récente, portant sur des images, a montré, à travers deux expériences, que les jugements portés sur la mémorabilité des images ne prédisaient pas leur mémorabilité *réelle*, telle que mesurée subséquemment par un test de mémoire (Isola et al., 2014). Nous avons trouvé une corrélation significative entre nos scores de mémorabilité et ceux de Libkuman *et al.*. Ce résultat concorde avec la littérature sur la méta-mémoire mais est incompatible avec celui d'Isola *et al.*. Il suggère qu'un individu est capable, dans une certaine mesure, de prédire sa mémorabilité réelle simplement en la regardant. Par conséquent, des scores de mémorabilité d'images obtenus par une annotation manuelle ne serait pas complètement décorrélés de scores de mémorabilité obtenus à l'aide d'une tâche de reconnaissance. Cette corrélation paraît cependant trop faible pour que l'annotation manuelle des images constitue une alternative sérieuse aux tests de mémoire pour constituer des bases de données pour l'étude de la mémorabilité des images.

Nous avons également observé une association entre les scores d'arousal des images et la différence entre la mémorabilité réelle des images et les jugements portés sur leur mémorabilité : plus l'arousal était haut, plus les participants tendaient à juger précisément de la mémorabilité des images. Ce résultat suggère que l'arousal est un indice valable pour juger de la mémorabilité d'une image. L'impression que nous avons généralement, lorsque nous sommes exposés à des évènements hautement stimulateurs, que nous ne les oublierons pas facilement, n'est donc pas infondée. Notons que, suivant ce résultat, il sera globalement plus facile de prédire la mémorabilité réelle des images d'une base de données présentant de nombreuses images activatrices, comme

l'IAPS, que d'une base de données en contenant moins, ce qui est probablement le cas de celle de (Isola et al., 2011b). Cela pourrait expliquer pourquoi Isola *et al.* ont obtenu un résultat différent du nôtre.

Les images émotionnellement chargées sont mieux reconnues

Nos résultats montrent que plus les images sont émotionnellement chargées, mieux elles étaient reconnues. Ils sont cohérents avec la littérature antérieure, présentée dans le chapitre 3 : les images — ou plus généralement, les stimuli — qui suscitent des émotions tendent à être mieux reconnues que les stimuli émotionnellement neutres (Kensinger and Schacter, 2008).

Un débat porte actuellement sur l'effet de la valence d'une information sur la probabilité qu'elle soit plus tard rappelée. Parfois, l'amélioration de la mémoire est comparable pour les stimuli positifs et négatifs (p. ex. (Bradley et al., 1992, Kensinger et al., 2002)). Parfois — en particulier dans les études portant sur la mémoire de stimuli verbaux ou d'images (Kensinger and Schacter, 2008) — les items négatifs sont plus susceptibles d'être rappelés que les items positifs (p. ex. (Charles et al., 2003, Ortony et al., 1983)). Et quelquefois — en particulier dans les études portant sur la mémoire d'évènements autobiographiques (Kensinger and Schacter, 2008) — les évènements positifs sont plus susceptibles d'être rappelés que les évènements négatifs (p. ex. (Argembeau et al., 2005, White, 2002)). Dans notre étude, les images négatives étaient globalement mieux reconnues que les images positives.

Kensinger *et al.* (Kensinger and Schacter, 2008) ont suggéré que les résultats conflictuels concernant l'effet de la valence sur la probabilité de se souvenir d'une information pouvaient être expliqués par la proposition selon laquelle les mécanismes de mémoire ont évolué pour faciliter l'encodage et la récupération de l'information affective la plus pertinente pour un but donné (thèse soutenue notamment par (Lazarus, 1991) et (Le Doux, 1996)). En effet, se remémorer un évènement négatif serait souvent utile pour la survie ou le bien-être, parce qu'en en refaisant l'expérience, nous serions plus à même d'en éviter la ré-occurrence (Le Doux, 1996). Aussi porterions-nous naturellement une attention accrue aux items négatifs, favorisant la mémorisation de ce type de matériel. Il y aurait également des situations dans lesquelles les évènements positifs seraient aussi pertinents, ou plus pertinents pour un but donné que les items négatifs. Par exemple, il a été montré que lorsque des stimuli positifs et négatifs étaient également liés aux préoccupations actuelles d'un individus, ils attiraient l'attention aussi bien l'un que l'autre (Riemann and McNally, 1995). En somme, il est possible que le but des participants à une tâche, qui varient selon les études, explique les différences observées entre celles-ci. Dans le cadre de la mémorisation intentionnelle d'images, dans lequel se sont inscrites l'ensemble des études sur la mémorabilité des images en vision par ordinateur, nos résultats confirment ceux obtenus par la plupart des études ayant porté sur la mémoire d'images : que les images négatives tendent à être plus mémorables que

les images positives.

L'écoulement du temps affecte la mémorabilité des images

L'analyse des résultats a montré que le degré de mémorabilité des images n'était plus le même un jour après l'encodage que quelques minutes après. D'autre part, les images les plus mémorables au premier test de mémoire n'étaient pas forcément les images plus mémorables au second test. Ces résultats signifient que la mémoire à long terme des images n'est pas stable. La mémoire des images est fragile, et sujette à des modifications, lors des premières 24 heures qui suivent l'encodage. Par conséquent, mesurer la mémorabilité d'images quelques minutes après leur encodage ne permet pas de dire que ces images seront durablement mémorables.

La théorie de la consolidation des souvenirs explique ce résultat. Selon cette théorie, les traces mnésiques sont soit renforcées, soit affaiblies (Brosch et al., 2013). Nos résultats renforcent l'hypothèse d'un rôle essentiel joué par l'arousal dans ce processus (McGaugh, 2000). En effet, ainsi que nous l'attendions, nous avons observé une association linéaire négative entre l'arousal et la différence de mémorabilité entre le premier test de mémoire et le second. Ce résultat suggère que l'arousal tend à préserver la mémoire de l'oubli. Il est cohérent avec la littérature antérieure (p. ex. (Bradley et al., 1992, Dolcos et al., 2004, Kensinger and Schacter, 2006)).

L'explication commune fait intervenir l'amygdale, une structure cérébrale en forme d'amande étroitement connectée à l'hippocampe (une autre structure du lobe temporal qui joue un rôle essentiel dans la mémoire à long terme (Zola-Morgan et al., 1986)), qui s'active lorsque qu'un individu ressent de l'arousal émotionnel, et qui jouerait un rôle majeur dans la consolidation mnésique (McGaugh, 2000). Suivant cette hypothèse, l'effet de l'arousal sur la mémoire lors de la phase de stockage s'explique (schématiquement) par la modulation de l'hippocampe par l'amygdale, qui conduit à la création de traces mnésiques plus robustes (Dolcos et al., 2004). En accord avec cette hypothèse, il a été montré que l'amygdale pouvait moduler les circuits neuronaux par lesquels s'exercent les processus de mémoire dans des situations où une émotion est ressentie (Brosch et al., 2013). Un autre argument de poids étaie la thèse d'un rôle de l'amygdale dans l'amélioration de la mémoire par l'arousal : les patients présentant des dommages à l'hippocampe ne montrent pas d'amélioration de la mémoire pour l'information émotionnelle (voir la revue de (Kensinger and Schacter, 2008)). Cependant, un tel résultat — s'il constitue une preuve considérable de la nécessaire contribution de l'amygdale dans l'augmentation de la mémoire par l'émotion — ne nous permet pas de savoir à quel stade de la mémoire (encodage, stockage ou récupération) l'amygdale exerce son influence. Des preuves plus directes nous viennent des études sur les animaux, qui ont montré que l'amygdale modulait l'activité de l'hippocampe durant la consolidation de la mémoire (McGaugh, 2000, McGaugh and Roozendaal, 2002) ; conclusion à laquelle ont également abouti quelques études, plus récentes, sur des humains (pour une revue

sur le sujet, voir (Kensinger, 2010)).

L'analyse de nos résultats a également mis en évidence une corrélation entre les scores de valence et la baisse de mémorabilité des images entre les deux tests de mémoire pour les images négatives, mais pas pour les images positives. Ce résultat signifie que la négativité des images tendaient à les préserver de l'oubli, mais par leur positivité. Cela suggère que la valence négative, mais pas la valence positive (ou dans une moindre mesure), joue un rôle dans la consolidation de la mémoire. Toutefois, la non-indépendance des dimensions d'arousal et de valence, avec une dissymétrie de la répartition des scores d'arousal en faveur de la valence négative (voir la section 6.2 du chapitre 6), doit nous faire considérer avec prudence cette interprétation.

Alors que de nombreux résultats suggèrent un rôle important joué par l'arousal (indépendamment de la valence) sur la rétention mnésique, le rôle joué par la valence (indépendamment de l'arousal) dans le cadre de la modulation de la mémoire par l'émotion est considéré comme secondaire (Dolcos et al., 2004, Kensinger and Schacter, 2006). Kensinger et Corkin ont montré l'existence de processus cérébraux distincts sous-jacents à l'amélioration de la mémoire pour l'information suscitant de l'arousal et pour l'information suscitant de la valence mais pas d'arousal (Kensinger and Corkin, 2004) : l'effet de l'arousal sur la mémoire serait dû à l'interaction entre l'amygdale et l'hippocampe, alors que l'effet de la valence serait dû à une interaction entre l'hippocampe et le cortex via l'activation d'un réseau neuronal impliqué dans les processus d'encodage contrôlé (en particulier l'élaboration). Ce résultat peut fournir une explication du rôle prépondérant de l'arousal dans la consolidation de la mémoire par rapport à la valence : contrairement à l'arousal, la valence pourrait n'avoir d'effet sur la mémoire que durant l'encodage mnésique, et pas pendant la phase de stockage.

7.5 Conclusion

Plusieurs résultats, qui ont des implications pour l'étude de la mémorabilité des images, ont été rapportés et discutés dans ce chapitre.

Premièrement, nos résultats confirment l'intérêt d'ouvrir l'étude de la mémorabilité des images à l'émotion qu'elles véhiculent. L'émotion véhiculée par les images est étroitement liée à leur mémorabilité. Par conséquent, il pourrait être intéressant d'utiliser cette information dans un objectif de prédiction de la mémorabilité des images. D'autre part, la répartition des images d'une base de données dans l'espace émotionnel arousal-valence est donc susceptible de biaiser les modèles prédictifs entraînés sur ces images (nous le confirmerons dans la section 8.4 du chapitre 8).

Deuxièmement, la mémorabilité des images baisse significativement durant le jour suivant leur encodage. Plus important encore, les images les plus mémorables quelques minutes après leur encodage ne sont pas nécessairement les images les plus mémorables

un jour après. Nous espérons que cette constatation encouragera les chercheurs en mémorabilité d'images à constituer des bases de données dont les images auront des scores de mémorabilité mesurés après des délais de rétention mnésique plus conséquents que dans (Isola et al., 2011b). En ajoutant qu'en obtenant des scores de mémorabilité mesurés après des délais de rétention mnésique variés pour un nombre conséquent d'images, il deviendra possible de lancer un nouveau champ de recherche, portant sur l'étude des caractéristiques qui font qu'une image est préservée de l'oubli.

Finalement, la comparaison de nos scores de mémorabilité avec ceux rapportés dans des études antérieures a jeter de la lumière sur deux points importants. D'une part, les modalité d'une tâche de reconnaissance ont une influence non négligeable sur les scores de mémorabilité obtenus. D'autre part, l'annotation manuelle des images paraît peu viable pour remplacer des mesures objectives afin d'obtenir des scores de mémorabilité.

Pris ensemble, ces différents résultats doivent attirer notre attention sur l'importance des données utilisées pour nourrir les modèles informatiques. Que cherche-t-on à prédire ? Et quel est l'écart entre ce qu'on cherche à prédire et ce que l'on prédit réellement ? La réponse à ces questions est toujours complexe.

Conclusion

Nous avons élaboré une nouvelle base de données — qui vient s'ajouter à l'unique base actuellement disponible (Isola et al., 2011b), à notre connaissance — pour l'étude de la mémorabilité des images. Les scores d'émotion et de mémorabilité que nous avons obtenus pour 150 images ont été comparés entre eux, et avec ceux obtenus dans des études antérieures. Nos analyses ont montré la cohérence des scores de mémorabilité et d'émotion que nous avons obtenus, au regard de ceux obtenus dans les études antérieures, de la théorie exposée dans la première partie de cette thèse, ainsi qu'en matière de cohérence inter-individuelle. La méthode proposée pourra être réutilisée pour obtenir des scores pour de nouvelles images afin d'obtenir un stock de données plus conséquent pour l'apprentissage automatique.

Les analyses menées ont confirmé deux points qui nous étaient apparus essentiels suite à notre étude de la littérature. D'une part, il est important d'ouvrir l'étude de la mémorabilité des images aux émotions qu'elles véhiculent. Nous espérons que notre base de données encouragera une telle ouverture. D'autre part, si notre communauté veut prédire une mémorabilité pérenne, il est important d'utiliser des scores mémorabilité obtenus à partir de performances de mémoire mesurées après un délai de rétention mnésique suffisant, ou de modéliser de manière suffisamment précise les liens entre de tels scores et ceux calculés à partir de performances de mémoire mesurées quelques minutes après l'encodage, pour inférer les uns des autres. Des scores de mémorabilité correspondant à des performances de mémoire mesurées après des durées de rétention différentes sont adaptés à un tel objectif ; ils permettent également d'étudier ce qui fait qu'une image n'est pas oubliée.

Dans la partie suivante, nous proposons une approche triple pour la prédiction de la mémorabilité des images, qui tienne compte à la fois des informations intrinsèques et extrinsèques (contextuelles et individuelles) des images.

III

Une approche triple pour répondre au défi de la prédiction de la mémorabilité

Introduction

Dans la partie précédente, nous avons mis en place une base de données constituée d'images avec des scores de mémorabilité, d'arousal et de valence. Des données sur les participants ont également été collectées. Nous mettons à profit ces différentes informations dans cette partie, où nous proposons une approche triple pour répondre au défi de la prédiction de la mémorabilité des images, qui repose : (1) sur l'utilisation de caractéristiques intrinsèques de l'image grâce à l'apprentissage profond, (2) sur la prise en compte du contexte de présentation de l'image, et (3) sur la prise en compte de l'observateur particulier de l'image.

Dans le chapitre 8, nous introduisons l'apprentissage profond (ou *Deep learning*) dans la prédiction de la mémorabilité des images. Cette technique puissante nécessitant un grand nombre d'exemples d'apprentissage, nous utilisons la méthode de l'ajustement fin (ou *Fine tuning*) pour l'apprentissage. Cette méthode consiste à ajuster un réseau de neurones pré-entraîné sur une base de données de large envergure (dans notre cas, destinée à la classification des images) à un problème donnée en poursuivant l'apprentissage sur une base de données de plus petite taille (dans notre cas, la base de (Isola et al., 2011b)).

Dans le chapitre 9, nous investiguons les effets du contexte de présentation d'une image sur sa mémorabilité, qui depuis récemment commencent à être pris en compte dans le cadre de la prédiction de la mémorabilité des images (Bylinskii et al., 2015b). En particulier, nous proposons de prendre en compte le contexte émotionnel de présentation des images. L'idée est d'améliorer les prédictions des modèles en utilisant les informations contextuels qui impactent la mémorabilité des images.

Dans le chapitre 10, nous nous intéressons, pour la première fois dans ce champ de recherche, à la prise en compte des facteurs individuels. Nous cherchons une voie qui permette d'intégrer l'idiosyncrasie dans l'équation. En effet, nous pensons que c'est à cette condition que nous parviendrons un jour à prédire avec précision la mémorabilité des images.

L'approche triple que nous proposons dans cette partie se veut un nouveau cadre de travail, plus large, pour la recherche sur la prédiction computationnelle de la mémorabilité des images, qui jusque-là s'est presque uniquement cantonnée aux caractéristiques intrinsèques des images.

8

Apprentissage profond pour la prédiction de la mémorabilité d'images

Les premières tentatives de prédiction de la mémorabilité d'images ont montré qu'il était possible d'extraire computationnellement d'une image des propriétés de bas niveau liées à sa mémorabilité, et d'en inférer – avec quelque succès – la probabilité qu'elle soit mémorisée. Dans ce chapitre, nous introduisons l'apprentissage profond pour la prédiction de la mémorabilité des images, et proposons un nouveau modèle prédictif : MemoNet. Nous éprouvons ensuite la capacité de prédiction de ce modèle sur notre base de 150 images, en comparant ses prédictions avec nos scores de mémorabilité comme vérité terrain. Nous étudions finalement les fluctuations de la performance de MemoNet en fonction du déplacement des images dans l'espace émotionnel arousal-valence, en nous appuyant sur les scores d'émotion que nous avons collectés pour nos 150 images. L'objectif est de mettre en lumière un éventuel biais dans notre modèle, entraîné sur la base de données de (Isola et al., 2011b), dont la répartition des images dans l'espace émotionnel n'a pas été évaluée.

8.1 Introduction

L'étude de la mémorabilité des images a récemment attiré l'attention des chercheurs en vision par ordinateur (Isola et al., 2011b, Bainbridge et al., 2013, Mancas and Le Meur, 2013). Les travaux existants ont montré que les individus partageaient une tendance à se rappeler et à oublier les mêmes images (Isola et al., 2014), ce qui ouvre la voie à

la conception de systèmes pour prédire la mémorabilité des images à partir d'informations qui leurs sont intrinsèques. La première tentative d'une telle prédiction a reposé sur l'utilisation de caractéristiques de bas niveau extraites des images, pour certaines manuellement, afin d'en prédire le degré de mémorabilité (Isola et al., 2011b). Les résultats montrent que la prédiction de la mémorabilité des images est possible uniquement à partir de leurs caractéristiques intrinsèques. Cependant, cette approche est modérément performante. L'apprentissage profond n'a, à notre connaissance, pas encore été essayé pour la prédiction de la mémorabilité des images. Au regard des récents résultats obtenus à l'aide de cette technique dans nombre de domaines, son utilisation pour la prédiction de la mémorabilité des images pourrait augmenter substantiellement la performance des modèles.

Les études qui ont porté sur la mémorabilité des images en vision par ordinateur (Isola et al., 2011b, Isola et al., 2011a, Khosla et al., 2012a, Khosla et al., 2012b, Mancas and Le Meur, 2013, Bylinskii et al., 2015a, Khosla et al., 2013, Isola et al., 2014, Celikkale et al., 2013, Celikkale et al., 2015, Kim et al., 2013), ont toutes, à ce jour, utilisé des images dont les scores de mémorabilité ont été obtenus avec la méthode proposée dans (Isola et al., 2011b). La base de données proposée par ces auteurs pourrait comporter des biais. La disponibilité d'autres bases d'images associées à des scores de mémorabilité, à l'image de celle que nous avons constituée, est essentielle, en premier lieu pour évaluer la capacité de généralisation des modèles. L'évaluation de notre modèle de prédiction — qui a également appris sur la base d'images d'Isola *et al.* (la seule qui comprenne une quantité d'images suffisante) — sur notre base de 150 images nous permettra de satisfaire ce point.

Cela nous permettra également d'éprouver la capacité de prédiction de notre modèle pour les images suscitant des émotions variées, en utilisant les scores d'émotion que nous avons obtenus. L'émotion qu'une image suscite est en effet, comme nous l'avons montré, étroitement liée à sa mémorabilité. Les images émotionnellement chargées sont généralement plus mémorables que les images émotionnellement neutres (Bradley et al., 1992, Cahill and McGaugh, 1995, Kensinger et al., 2002). Or, la distribution dans l'espace émotionnel des images utilisées pour entraîner et éprouver les modèles de prédiction de la mémorabilité existants n'a pas été évaluée. Par conséquent, il est possible qu'une répartition particulière de ces images dans l'espace émotionnel ait entraîné un biais *émotionnel* dans les modèles de prédiction nourrit avec ces images. En testant le modèle à apprentissage profond que nous introduisons dans ce chapitre, MemoNet, dont l'apprentissage a été réalisé à l'aide des images de la base de (Isola et al., 2011b), sur notre base de données, nous pourrons nous rendre compte si sa performance de prédiction fluctue en fonction de la charge émotionnelle des images. Le cas échéant, cela montrerait l'importance des émotions véhiculées par les images utilisées pour entraîner les modèle de prédiction de la mémorabilité, et constituerait un argument en faveur de l'ouverture de notre champ de recherche à leur prise en compte.

8.2 Contexte

Isola *et al.* ont été les premiers, à notre connaissance, à chercher à prédire computationnellement la mémorabilité d'images. Dans ce but, ils ont déterminé une combinaison de caractéristiques globales de l'image pour en prédire la mémorabilité en utilisant un séparateur à vaste marge pour la régression (SVR) (Isola et al., 2011b, Isola et al., 2014). Les caractéristiques utilisées ont été fréquemment utilisées dans des travaux récents de vision par ordinateur, en particulier dans un objectif de reconnaissance de scènes et d'objets (voir le chapitre 4) : les histogrammes des couleurs et de gradient orienté (HOG2x2), les descripteurs SIFT (pour *Scale-Invariant Feature Transform*) et GIST, et l'indice de similarité structurelle (SSIM).

Plus récemment, Mancas et Le Meur ont montré que l'utilisation de caractéristiques liées à l'attention visuelle remplaçait avantageusement certaines caractéristiques de bas niveau utilisées par (Isola et al., 2011b), en atteignant une performance de prédiction très légèrement meilleure tout en ayant réduit considérablement le nombre de caractéristiques d'entrée (Mancas and Le Meur, 2013). Plus précisément, l'utilisation de caractéristiques liées à l'attention visuelle a permis une augmentation de performance de 2% en utilisant 17 dimensions au lieu des 512 dimensions du descripteur GIST, en plus des autres caractéristiques introduites dans (Isola et al., 2011b) (i.e. SIFT, HOG2x2, SSIM, et les histogramme des couleurs).

Les modèles existants reposent donc sur des combinaisons prédéfinies de caractéristiques dites *handcrafted*. Or, définir ce type de caractéristiques est coûteux et demande une connaissance approfondie du domaine étudié. Contrairement aux travaux précédents, qui reposent sur l'utilisation de caractéristiques *handcrafted*, nous présentons dans ce chapitre une étude dans laquelle nous utilisons un réseau de neurones convolutifs (CNN, pour *Convolutional Neural Network*) pour prédire le degré de mémorabilité d'images. Les CNNs sont composés de couches de convolution empilées, suivies par une ou plusieurs couches entièrement connectées (LeCun et al., 1989). L'approche basée sur l'utilisation de CNN a eu un impact considérable dans le domaine de l'apprentissage automatique. À ce jour, les meilleurs résultats apportés pour répondre au défi de classification d'ImageNet[1] ont été obtenus en utilisant des modèles basés sur des CNN (Krizhevsky et al., 2012, Szegedy et al., 2015).

Cependant, de larges bases de données étiquetées sont nécessaires pour entraîner les CNN, et il n'existe pas actuellement de bases assez larges d'images associées à des scores de mémorabilité. Néanmoins, la stratégie de l'ajustement fin (*fine-tuning*) s'avère efficace pour l'apprentissage de CNN quand les données sont rares (Girshick et al., 2014). Cette stratégie d'ajustement fin consiste à pré-entraîner un CNN sur une base de données de large envergure mais pas adaptée à la question à traiter (dans notre

[1] ImageNet est une base de données de plusieurs millions d'images destinée à la recherche sur les algorithmes de reconnaissance visuelle d'objets (Deng et al., 2009). Les images ont été annotées par des annotateurs humains, dont la tâche était de désigner les objets représentés dans les images.

cas, ImageNet, destinée à la classification des images), puis d'ajuster le réseau pré-entraîné en continuant la rétro-propagation sur une base de données de plus petite taille mais adaptée au problème à résoudre (dans notre cas la base de données de (Isola et al., 2011b)) pour qu'il puisse résoudre la tâche de classification désirée. Le principe de cette stratégie repose sur le fait que les caractéristiques les plus globales d'un CNN sont contenues dans ses premières couches et peuvent être utilisées pour résoudre des tâches variées (alors que les dernières couches du réseau sont de plus en plus spécifiques et n'ont d'intérêt que pour la tâche assignée au CNN).

8.3 Prédiction de la mémorabilité

Les modèles pré-entraînés à ajustement fin destinés originellement à l'extraction de la sémantique de la scène et des objets dans les images nous ont semblé particulièrement adaptés à la prédiction de la mémorabilité, comme ces deux concepts sont intrinsèquement liés (Isola et al., 2014). Nombre des caractéristiques *handcrafted* utilisées par Isola *et al.* pour la prédiction de la mémorabilité ont d'ailleurs été précédemment utilisées pour l'extraction de la sémantique de la scène et des objets. Les réseaux de neurones biologiquement inspirés ont également modélisé avec quelque succès les processus de mémoire (Abraham and Robins, 2005, Ans and Rousset, 2000, McClelland et al., 1995). Pour ces raisons, nous avons utilisé — et ajusté à la prédiction de mémorabilité — le modèle GoogleNet, introduit par Szegedy *et al.*, qui est depuis 2014 la référence en matière de performance de catégorisation et de détection sur la base de données ImageNet (Szegedy et al., 2015).

8.3.1 Réseau de neurones convolutifs à ajustement fin

L'architecture de GoogleNet (schématiquement représentée dans la figure 8.1) est un assemblage de neuf réseaux similaires, dénommés « modules d'inception » (Szegedy et al., 2015). Un module ou réseau d'inception se compose d'un empilage de couches de convolution 1×1, 3×3, et 5×5, ainsi que de couches de *max-pooling* insérées périodiquement entre deux couches convolutives successives, destinées à réduire la quantité de calcul et à éviter le sur-apprentissage. Le pooling ou « mise en commun » consiste à sous-échantillonner l'image d'entrée, c'est-à-dire à la découper en plusieurs rectangles de n pixels de côté qui ne se chevauchent pas. Étant donnée la profondeur du réseau, deux couches de pertes auxiliaires, en plus de la couche de perte habituelle qui correspond à la dernière couche du réseau, ont été connectées aux couches intermédiaires pour augmenter la rétro-propagation du gradient et éviter le problème dit « d'évaporation »(ou *vanishing gradient problem*). Les couches de perte, représentées par des rectangles jaunes dans la Figure 8.1, spécifient comment l'entrainement du réseau pénalise l'écart entre le signal prévu et réel. Durant l'entraînement, les résultats des couches de

perte auxiliaires sont ajoutés à la perte finale du réseau, avec un poids moindre dans la somme cependant. L'entraînement prend fin dès qu'un certain nombre d'itérations est atteint. Au moment du test, les couches de perte auxiliaires n'étant plus nécessaires, elles sont retirées du réseau.

Dans notre approche par ajustement fin, nous avons remplacé les trois couches de perte (i.e. les deux couches de perte auxiliaire et la couche de perte finale) par une seule couche entièrement connectée, composée d'un unique neurone. Les fonctions de perte associées au modèle correspondent à des pertes euclidiennes. Toutes les couches du modèle pré-entraîné sont ajustées (*fine-tuned*), mais le taux d'apprentissage associé aux couches originelles est dix fois plus petit que celui associé au nouveau dernier neurone. L'objectif est que les couches pré-entraînées changent très lentement, tout en permettant à la nouvelle couche d'apprendre rapidement. Les poids de la nouvelle couche sont initialisés en utilisant l'algorithme Xavier, qui détermine automatiquement l'échelle d'initialisation à partir du nombre de neurones d'entrée et sortie (Glorot and Bengio, 2010). Le modèle à ajustement fin proposé pour la prédiction de la mémorabilité d'images sera appelé MemoNet dans la suite de ce chapitre.

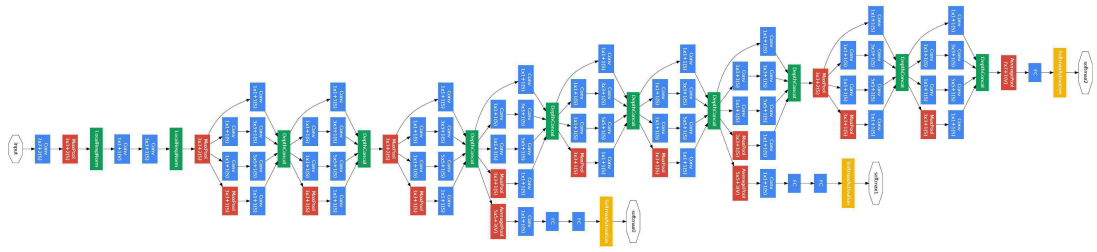


FIGURE 8.1 – Schéma de l'architecture de GoogleNet, introduit par Szegedy *et al.* (Szegedy et al., 2015).

8.3.2 Résultats obtenus par MemoNet

Afin d'obtenir des résultats comparables avec les travaux précédents, nous avons utilisé les mêmes données et protocoles d'entraînement et de test qu'Isola *et al.* (Isola et al., 2014) et que Mancas et Le Meur (Mancas and Le Meur, 2013). La base de données, présentées dans la première partie de ce mémoire, est composée de 2222 images associées à des scores de mémorabilité obtenus à l'aide d'un test de mémoire présenté sous la forme d'un jeu en crowdsourcing (Isola et al., 2011b). Les images notées ont été sélectionnées aléatoirement dans la base de données SUN (Xiao et al., 2010) et représentent des types de scène variés. La catégorie de la scène représentée et celles des objets à l'intérieur de la scène sont disponibles pour chaque image (voir (Choi et al., 2010)). Un score de mémorabilité pour une image, utilisé comme "vérité terrain" pour entraîner et tester MemoNet, est défini comme le pourcentage de détections correctes

de l'unique répétition de l'image dans un flux d'image quelques minutes après sa présentation initiale, calculé sur un ensemble de participants. MemoNet a été entraîné 25 fois en utilisant les ensembles pour l'apprentissage définis par Isola *et al.*, composés de la moitié des images, et testé sur l'autre moitié des images.

Comme pour Isola *et al.* (2011), la performance globale est définie comme le coefficient de corrélation des rangs de Spearman (ρ) moyen correspondant à la moyenne des coefficients de Spearman entre les scores de mémorabilité correspondant à la vérité terrain et les prédictions de MemoNet obtenus pour chacun des 25 modèles entraînés. Nous avons également calculé l'erreur quadratique moyenne (notée MSE pour *Mean Squared Error*). Les performances de MemoNet, obtenues après un certain nombre d'itérations (1k, 10k et 30k), ainsi que les performances des modèles obtenues dans les études précédentes, sont indiquées dans la Table 8.1.

Les résultats montrent que la performance de MemoNet dépasse significativement les performances obtenues par Isola *et al.* (Isola et al., 2014) et par Mancas et Le Meur (Mancas and Le Meur, 2013). En particulier, MemoNet 30k montre un gain de performance de 32.78% par rapport à Mancas et Le Meur, qui jusque-là (c'est-à-dire en 2015) avaient obtenus la meilleure performance de prédiction.

D'autre part, la performance de prédiction de MemoNet peut être rapprochée de celles des humains. Dans la section 7.1.2 du chapitre 7, nous avons observé une corrélation de Pearson de .17 entre les scores de mémorabilité de (Libkuman et al., 2007), qui correspondent à des jugements portés a priori sur la mémorabilité des images, et ceux de notre base de données, qui correspondent à une mesure objective de la mémoire. (La corrélation de Spearman est également de $\rho = .17, p < .05$.) Pris ensemble, ces résultats suggèrent que la performance de prédiction de MemoNet est bien supérieure à celle d'un être humain.

TABLE 8.1 – Performances de prédiction de la mémorabilité d'images globales mesurées par le coefficient de corrélation des rangs de Spearman (ρ)moyen, et les erreurs quadratiques (MSE) associées.

	ρ	MSE
Isola *et al.*, 2014	0.462	0.017
Mancas et Le Meur, 2013	0.479	X
MemoNet 1k	0.522	0.017
MemoNet 10k	0.620	0.012
MemoNet 30k	0.636	0.012

La Table 8.2 présente la performance de prédiction de MemoNet 30k en fonction des différentes catégories de scène et d'objet comprenant au moins 100 images. Il est intéressant de noter que, comme pour le modèle d'Isola *et al.* (2014), MemoNet 30k

obtient ses meilleures performances dans les catégorie "Personne(s)" et "Personne(s) assise(s)". Nous n'avons trouvé aucune corrélation significative entre la taille de la catégorie d'objets et la performance de MemoNet 30k pour la catégorie ($r = -0.018, t(25) = -0.092, p = .46$) ; une telle corrélation (positive) aurait signifié que notre réseau obtenait de meilleures performances pour les catégories pour lesquelles le nombre d'images ayant servi à son entraînement était plus important.

8.4 La performance de MemoNet sur notre base de données

Les modèles existants de prédiction de la mémorabilité (Isola et al., 2011a, Isola et al., 2014, Mancas and Le Meur, 2013) — le notre compris — ont été mis au point en utilisant la même base d'images, la base d'Isola *et al.* (Isola et al., 2011b). D'autre part, l'émotion véhiculée par une image est un facteur clé de sa mémorabilité (p. ex. (Abrisqueta-Gomez et al., 2002, Bradley et al., 1992). Pour ces deux raisons, nous avons testé la performance de MemoNet sur la base de 150 images que nous avons constituée afin de déterminer la capacité de notre modèle à généraliser sur une nouvelle base d'images associées à des scores de mémorabilité (la seule existant à ce jour), et d'étudier si sa performance variait avec l'émotion véhiculée par les images (grâce aux scores d'émotion que nous avons également obtenus).

Nous ne considèrerons que les scores de mémorabilité obtenus au premier test de mémoire, obtenus avec une méthode proche de celle utilisée par (Isola et al., 2011b). Pour nous rendre compte si la performance de notre modèle, entraîné (en partie) sur la base d'Isola *et al.* (2011), est identique quel que soit l'émotion suscitée par l'image donnée en entrée, nous utiliserons les scores d'arousal et de valence que nous avons obtenus pour nos 150 images, ainsi que les scores d'arousal, de valence, et de dominance obtenus par (Lang et al., 2008) et (Ito et al., 1998) pour ces mêmes images tirées de l'IAPS. La dominance était, à l'instar de la valence et de l'arousal, mesurée par une échelle SAM en neuf points, allant d'un état mental " dominé " par l'émotion induite (1) à un état de " contrôle " (9).

Nous avons pré-traité nos 150 images afin de les mettre au format des images ayant servi d'entrées pendant l'apprentissage de MemoNet (i.e. des images redimensionnées en 224 $\times$ 224, coupées si nécessaire à partir du centre pour éviter la déformation). Avant de couper nos images, nous avons supprimé les bandes noires mises sur les côtés de certaines images par Lang *et al.* pour obtenir un unique format pour l'ensemble des images de l'IAPS (Lang et al., 1997, Lang et al., 1999, Lang et al., 2008).

La performance globale de MemoNet 30k pour cette nouvelle base d'images est plus faible que celle obtenue sur les images de la base d'Isola *et al.* ($\rho = 0.251; MSE = 0.033$).

TABLE 8.2 – Influence de la catégorie de la scène et/ou des objets (pour les catégories comprenant au moins 100 exemplaires) sur la performance de prédiction de MemoNet 30k. Le nombre d'images dans la catégorie et la moyenne des scores de mémorabilité (notée VT, pour Vérité Terrain) des images de la catégorie sont également indiqués.

Rang	**Catégorie**	**Taille**	**VT**	ρ
1	person sitting	165	0.753	0.655
2	person	554	0.725	0.628
3	pole	108	0.667	0.613
4	mountain	272	0.593	0.606
5	painting	101	0.696	0.593
6	wall	989	0.718	0.581
7	window	589	0.662	0.580
8	table	212	0.703	0.580
9	sign	147	0.664	0.572
10	door	361	0.668	0.570
11	chair	268	0.712	0.564
12	fence	181	0.656	0.554
13	sky	1080	0.628	0.550
14	tree	814	0.630	0.548
15	plant	417	0.640	0.543
16	floor	766	0.727	0.537
17	ground	269	0.637	0.521
18	ceiling	571	0.713	0.514
19	water	151	0.631	0.511
20	road	297	0.647	0.497
21	sidewalk	163	0.643	0.493
22	building	699	0.630	0.492
23	ceiling lamp	289	0.713	0.491
24	box	121	0.719	0.489
25	steps	115	0.659	0.477
26	grass	341	0.630	0.464
27	car	192	0.649	0.464

TABLE 8.3 – Coefficient de corrélation de rangs de Spearman (ρ) entre les performances locales de MemoNet 30k pour les 150 images de notre base et leurs scores d'émotion ($^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$)

Dimension	Origine des données	ρ
Valence	Lang *et al.* (Lang et al., 2008)	-0.285***
	Ito *et al.* (Ito et al., 1998)	-0.248**
	Notre base de données	-0.284***
Arousal	Lang *et al.* (Lang et al., 2008)	0.096
	Ito *et al.* (Ito et al., 1998)	0.222*
	Notre base de données	0.198**
Dominance	Lang *et al.* (Lang et al., 2008)	-0.221**

Une performance locale (i.e. pour une image) de MemoNet 30k est définie comme la différence entre le score de mémorabilité de l'image correspondant à la vérité terrain et la moyenne des prédictions de mémorabilité pour l'image par les 25 modèles entraînés. La table 8.3 présente les corrélations entre les performances de MemoNet 30k et les scores d'émotion des 150 images de la base utilisée, obtenus dans notre étude et dans les études précédentes de (Lang et al., 2008) et (Ito et al., 1998). Ces corrélations mesurent la relation entre la dimension émotionnelle considérée (arousal, valence ou dominance) et l'écart de prédiction de mémorabilité moyen de notre modèle à la vérité terrain — c'est-à-dire que le modèle prédise un score de mémorabilité supérieur à la vérité terrain, ou inférieur. Il faut également noter que les performances locales pour les données d'Ito *et al.* sont évaluées à partir d'un sous-ensemble d'images, comme ces auteurs n'ont fait évaluer qu'une partie des images de l'IAPS utilisées dans notre base de données (104 images sur les 150 que nous avons utilisées). Les performances locales de MemoNet 30k et les scores d'émotion que nous avons obtenus (dans notre étude présentée dans le chapitre 5) sont présentées dans la figure 8.2.

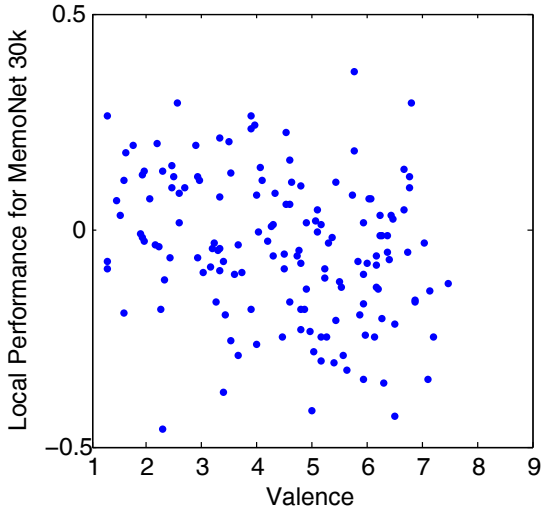
Les corrélations de rangs montrent une relation de force modérée, mais cohérente, entre la performance de MemoNet 30k et les scores de valence, d'arousal, et également de dominance, pour les scores d'émotion de notre étude et des études de (Ito et al., 1998) et de (Lang et al., 2008) — à l'exception des scores d'arousal de (Lang et al., 2008). La valence est négativement corrélée avec la performance locale de MemoNet 30k, alors que l'arousal est positivement corrélé à celle-ci. Comme la valence, la dominance est négativement corrélée avec la performance locale de notre modèle. Cette dimension émotionnelle étant considérée comme secondaire, et étant souvent mal comprise par les annotateurs, nous la laisserons de côté dans la suite des analyses.

En vue d'analyser avec plus de profondeur la relation entre la performance de prédiction de notre modèle et l'émotion véhiculée par les images, nous avons utilisé un algo-

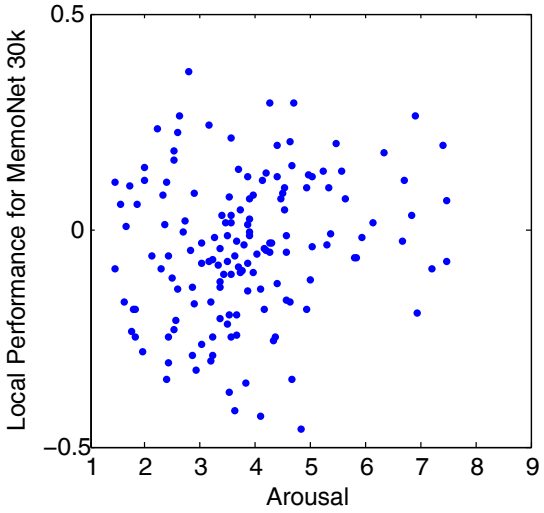
rithme de partitionnement en *k*-moyennes pour séparer les 150 images en trois groupes sur la base de leurs positionnements dans l'espace bidimensionnel valence-arousal (voir la figure 8.2(c)). Approximativement, les trois groupes séparent les images qui véhiculent des émotions négatives et activatrices (groupe 1) des images qui véhiculent des émotions neutres (groupe 2) et des images qui véhiculent des émotions positives et modérément activatrices (groupe 3). Il est important de noter que les dimensions d'arousal et de valence sont corrélées entre elles, ainsi que nous l'avons montré dans la section 6.2 du chapitre 6.

Une ANOVA à un facteur montre que le groupe d'images a un effet sur la performance locale de MemoNet 30k ($F(2, 147) = 5.82; p < .005$). Un test de comparaisons multiples de Tukey révèle que la performance locale moyenne de notre modèle pour les images du groupe 1 ($\mu = 0.0298$) est meilleure (i.e. plus proche de 0, soit l'absence d'écart entre les prédictions du modèle et la vérité terrain) que la performance pour les images du groupe 2 ($\mu = -0.0536$) et du groupe 3 ($\mu = -0.0847$); et que la performance locale moyenne de notre modèle est meilleure pour les images du groupe 2 que pour les images du groupe 3. En d'autres termes, MemoNet 30k est meilleur pour prédire la mémorabilité des images négatives et activatrices que des images neutres et positives, moins activatrices. Il y a donc un biais dans notre modèle, qui suggère que, l'émotion et la mémorabilité étant liées, la performance de prédiction de mémorabilité de notre modèle a dépendu de l'émotion véhiculée par les images de la base utilisée pour entraîner le modèle. On peut en inférer l'existence d'un biais dans la base de données d'Isola *et al.*, sur laquelle notre modèle a appris lors de l'ajustement fin. Il se pourrait, par exemple, que cette base contiennent plus d'images négatives et activatrices que d'images neutres et positives — ce qui aura fait que notre modèle a été plus performant pour ces premiers exemplaires. Bien sûr, ce biais pourrait être dû à des interactions plus complexes. Quoiqu'il en soit, il apparaît important de s'intéresser à la répartition des images d'une base de données destinée à la prédiction de la mémorabilité dans l'espace des émotions (p. ex. dans l'espace arousal-valence, dont on a montré qu'il paraissait le plus pertinent pour appréhender l'émotion). En effet, comme nos résultats le suggèrent, même la sélection aléatoire d'images (p. ex. dans une base de données ou sur internet) ne garantit pas une répartition "naturelle" des images dans un tel espace. Encore faudrait-il définir ce qu'une répartition "naturelle" signifie : c'est là qu'on voit toute l'importance du contexte dans la mémorabilité des images.

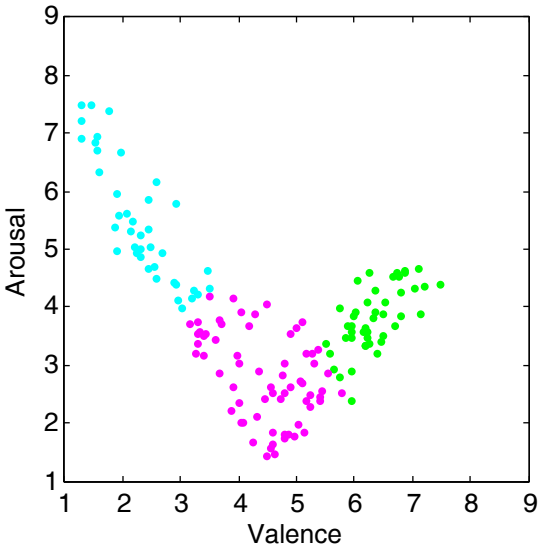
Les résultats suggèrent également que la prise en compte de l'émotion véhiculée par les images dans les modèles pourraient améliorer les performances de prédiction. En ceci, ils constituent donc un argument supplémentaire en faveur de l'ouverture de l'étude de la mémorabilité des images en vision par ordinateur aux émotions véhiculées par les images.



(a) Scores de valence



(b) Scores d'arousal



(c) Résultat du partitionnement en k-moyennes

FIGURE 8.2 – Performance locale de MemoNet 30k en fonction des scores (a) de valence et (b) d'arousal que nous avons obtenus pour les 150 images de notre base, et (c) le résultat du partitionnement en k-moyennes ($k = 3$).

8.5 Conclusion

Nous avons introduit l'apprentissage profond pour la prédiction de la mémorabilité des images. Ce modèle surpasse en performance les modèles existants, et montre, sur la même base de données et dans des conditions d'évaluation équivalentes, une augmentation de performance de prédiction de 32.78% par rapport au plus performant des modèles de l'état de l'art. Ce résultat démontre l'intérêt d'utiliser l'apprentissage profond dans l'étude de la mémorabilité des images, et la viabilité de la méthode d'ajustement fin utilisée pour palier la faible quantité d'images (par rapport au besoin, important, de l'apprentissage profond) évaluées selon leur degré de mémorabilité.

Cependant, la généralisation de la performance de notre modèle à une nouvelle base de données (i.e. la nôtre) est un succès mitigé. Cela soulève le problème de la représentativité de l'unique base d'images notées sur la dimension de mémorabilité (Isola et al., 2011b) (à part nous), utilisée pour entraîner l'ensemble des modèles existants. Cela vient également rappeler l'importance d'avoir à disposition plusieurs bases de données lorsqu'on met au point des modèles de prédiction. Sur ce dernier point, la base que nous avons mis en place, et la méthode pour l'augmenter, constitue un pas dans cette direction, même si le nombre d'images, limité, ne permettrait pas à lui seul d'entraîner un CNN.

Il est apparu que notre modèle présentait un biais : il obtient de meilleures performances de prédiction pour les images négatives suscitant de l'arousal que pour les images neutres ou positives, qui suscitent un arousal moindre. Ce résultat renforce notre opinion, plusieurs fois évoquée dans ce mémoire, sur l'importance de la prise en compte de l'émotion véhiculée par les images dans une base visant à étudier leur mémorabilité. Il montre l'importance de la répartition des images d'une base destinée à l'étude de la mémorabilité dans l'espace des émotions. Pour avoir une idée de cette répartition, il est possible d'extraire automatiquement de l'information émotionnelle des images ; c'est en effet l'objet d'un champ de recherche particulier (p. ex. (Wei et al., 2008, Gbèhounou et al., 2012)), encore immature, mais qu'on gagnerait à lier à celui de la mémorabilité, d'autant qu'il progressera.

Le fait qu'une image suscitera une émotion particulière, ou qu'elle sera plus ou moins bien mémorisée, dépend en partie de l'observateur particulier de l'image et du contexte dans lequel il se trouve. Parce que la mémorabilité est subjective, sa prédiction est condamnée à l'imprécision si l'on ne s'intéresse qu'à l'information intrinsèque des images. Pour cette raison, dans les chapitres suivants, nous nous intéresserons à l'exploitation d'informations extrinsèques liées à la mémorabilité des images : les informations liées au contexte (chapitre 9) et à l'utilisateur (chapitre 10), avec l'objectif d'intégrer ces informations à la prédiction de la mémorabilité pour *personnaliser* la prédiction en corrigeant les scores obtenus par MemoNet.

9

Le contexte de présentation d'une image et sa mémorabilité

Il a été montré que l'extraction de diverses informations intrinsèques de l'image permettait de prédire dans une certaine mesure sa mémorabilité (Isola et al., 2014, Mancas and Le Meur, 2013). Cependant, la précision de la prédiction souffre de l'absence de prise en compte des facteurs extrinsèques, à la fois individuels et contextuels. Ce chapitre traite des effets contextuels sur la mémorabilité des images. En particulier, nous nous intéressons à l'effet de la fréquence d'occurrence de la catégorie sémantique d'appartenance de l'image et aux effets du contexte émotionnel de présentation des images. L'objectif recherché est de mettre en lumière des effets contextuels qui pourront être pris en compte automatiquement dans la prédiction de la mémorabilité des images.

9.1 Introduction

Jusqu'à très récemment (Bylinskii et al., 2015b), aucun des travaux portant sur la prédiction computationnelle de la mémorabilité des images n'avait, à notre connaissance, porté sur la prise en compte du contexte de présentation des images dans les modèles computationnels. En psychologie, l'effet du contexte sur la mémoire est portant étudié depuis des décennies, en particulier dans les études portant sur la distinctivité des items mémorisés (p. ex. (Eysenck, 2014, Nairne, 2006, Schmidt, 1985)) ou celles portant sur les effets contextuels de l'émotion sur la mémoire (p. ex. (Murray, 1999, Ucros, 1989)).

Bylinskii *et al.* ont proposé une première méthode pour prendre en compte auto-

matiquement le contexte de présentation d'une image dans les modèles de prédiction de la mémorabilité des images (Bylinskii et al., 2015b). À partir de caractéristiques extractibles de l'image, leur méthode permet de quantifier le degré par lequel les images se distinguent de leur contexte de présentation. Les auteurs montrent que plus l'image est distincte de son contexte, plus elle est mémorable. Le travail présenté dans la première section de ce chapitre (réalisé avant à la parution de cet article) s'inscrit dans une démarche similaire : nous cherchons à mettre en évidence un facteur contextuel — la fréquence d'occurrence de la scène représentée par une image dans un ensemble d'images — susceptible d'influencer la mémorabilité des images.

Le contexte émotionnel de présentation des images, dont la littérature en psychologie suggère qu'il influerait sur leur mémorabilité, n'a, à notre connaissance, jamais été pris en compte dans l'étude de la mémorabilité des images en vision par ordinateur. Or, il est possible, comme nous l'avons précédemment évoqué, d'extraire computationnellement, dans une certaine mesure, les émotions véhiculées par les images (Wei et al., 2008, Gbèhounou et al., 2012). Cette information émotionnel pourrait être utilisée pour prendre en compte automatiquement le contexte émotionnel de présentation des images pour en prédire la mémorabilité. Dans un tel objectif, il serait nécessaire de révéler préalablement l'existence d'effets contextuels de l'émotion sur la mémoire des images. Les travaux rapportés dans la seconde section de ce chapitre se veulent un premier pas dans cette direction.

9.1.1 Effet de la fréquence d'occurrence de la scène représentée par l'image

Comme nous l'avons expliqué dans la section 4.1.2 du chapitre 4, Isola *et al.* ont évalué l'effet du contexte dans leur étude, en mesurant la corrélation entre la fréquence du contenu des images dans la base de données utilisée et la mémorabilité moyenne des images (Isola et al., 2011b). Plus précisément, ces auteurs se sont intéressés à la fréquence dans leur base de données du type d'objet contenu dans les images et du type de scène représenté dans les images. Les corrélation de Spearman obtenues sont les suivantes : $\rho = -0.01$ et $\rho = -0.13$, respectivement. La dernière corrélation seulement est significative. Elle suggère qu'une scène moins représentée dans une base de données, dont on pourrait imaginer qu'elle gagne en distinctivité, est plus mémorable. Les auteurs se contentent d'expliquer qu'ils n'ont pas trouvé de corrélation forte, et que cela suggère que de simples formes de biais de fréquence n'expliquent pas les résultats de mémorabilité obtenus pour leurs images. Ils recommandent cependant, pour tester de plus subtiles interactions de la mémorabilité avec le contexte de présentation des images, de mesurer la mémorabilité sur de nouveaux jeux d'images, ce qui permettrait de mesurer la généralisabilité de leurs résultats.

Comme nous l'avons évoqué au début de cette introduction, Bylinskii *et al.* ont

proposé une méthode pour prendre en compte le contexte de manière automatique dans la prédiction de la mémorabilité des images (Bylinskii et al., 2015a). Ils ont calculé pour chaque image la mesure dans laquelle elle était distincte de son contexte sur la base d'un ensemble de caractéristiques extraites des images. En particulier, ils se sont intéressés au type de scène représenté dans l'image (p. ex. « Parc d'attraction », « cockpit »). Comme dans (Isola et al., 2011b, Isola et al., 2014), ils ont observé que certains types de scène sont plus mémorables que d'autres. Quant aux effets contextuels, ils ont montré un effet de la distinctivité du type de scène représenté dans une image par rapport aux types de scène représentés par les autres images (i.e. le contexte) sur la mémorabilité de cette image.

Le type de scène représenté dans l'image est donc un facteur à partir duquel peuvent être quantifiés certains effets contextuels sur la mémorabilité des images. Dans le cadre de la prédiction computationnelle de la mémorabilité des images, un modèle utilisera un algorithme de classification pour attribuer automatiquement à chaque image une catégorie de scène d'appartenance. Il existe aujourd'hui de très puissants algorithmes de ce type, par exemple (Zeiler and Fergus, 2014, Szegedy et al., 2015).

9.1.2 Effets contextuels de l'émotion sur la mémoire

Les effets de l'émotion sur la mémoire peuvent se produire à différentes étapes : lors de l'encodage, du stockage et/ou de la récupération mnésiques. Durant l'encodage, l'émotion peut prioriser la perception (Sharot and Phelps, 2004, Ochsner, 2000) et le traitement (Kensinger, 2004) de l'information émotionnellement chargée en influençant l'attention sélective et le temps de regard, ce qui favoriserait son encodage (et, par conséquent, sa récupération ultérieure). Lors du stockage, plusieurs études montrent que l'émotion — en particulier, l'arousal — influence la consolidation des souvenirs (Cahill and McGaugh, 1995, McGaugh, 1992). Nous nous sommes intéressés à ce point dans le chapitre 7.

Les effets contextuels de l'émotion sur la mémoire sont susceptibles d'advenir lors des phases d'encodage et de récupération des souvenirs. En particulier, deux effets ont fait l'objet d'études : l'effet de congruence émotionnelle (Blaney, 1986) et la récupération dépendante du degré de similarité entre l'état émotionnel lors de l'encodage et l'état émotionnel lors de la récupération (ou *Mood-state dependant retrieval* (Ucros, 1989). L'effet de congruence émotionnelle renvoie à la tendance des individus à récupérer plus facilement de l'information en mémoire lorsqu'elle est émotionnellement proche de celle de leur état émotionnel (Lewis and Critchley, 2003). Cet effet a été montré pour des états émotionnels invoqués et durables ; par exemple, les personnes dépressives récupèreraient en mémoire plus de souvenirs de valence négative que positive (Murray, 1999, Watkins et al., 1996). Il a également été montré pour des états émotionnels provoqués ; par exemple, l'induction en laboratoire d'une émotion positive par l'écoute d'une musique joyeuse augmenterait la probabilité de se remémorer des souvenirs joyeux de

son enfance (Martin and Metha, 1997). La récupération en mémoire d'une information est généralement plus efficace lorsque l'état émotionnel de l'individu au moment de la récupération est proche de son état émotionnel au moment où il a encodé l'information (Ucros, 1989). De tels effets sont, dans une tâche de reconnaissance d'images, susceptibles de modifier la probabilité de reconnaître certaines images (comme nous l'avons évoqué dans la section 3.1.3 du chapitre 3).

Ces deux effets se rapportent à la proximité entre l'humeur des individus, qui renvoie à un état affectif durable, et l'émotion suscitée par un stimulus, éphémère. Ils nous rappellent que la mémoire humaine a tendance à relier des événements avec des significations affectives similaires (Cahill and McGaugh, 1995). En extrapolant, on pourrait imaginer que des effets similaires se produisent lorsque, au lieu de l'humeur des participants, on considèrerait l'émotion suscitée par des images dans une tâche de reconnaissance. En pratique, on ne s'intéresserait plus à la proximité entre l'humeur d'un participant et la coloration émotionnelle d'une image cible, mais, par exemple, à la proximité entre la coloration émotionnelle d'une image qu'on pourrait appeler « amorce » et celle d'une image cible. En s'inspirant de l'effet de congruence émotionnelle, on pourrait alors imaginer que la proximité en matière d'émotion véhiculée entre une image amorce et une image cible pourrait faciliter la récupération de cette dernière. Cette influence irait donc dans le sens d'une augmentation de la probabilité de reconnaître une image cible lorsqu'elle induit chez l'observateur un état émotionnel proche de celui induit par l'image qui la précède. De même, en s'inspirant de l'effet de *Mood-state dependant retrieval*, on pourrait émettre l'hypothèse qu'une image cible a plus de chance d'être reconnue si le contexte émotionnel local de son encodage est proche du contexte émotionnel local de sa récupération (en entendant par contexte local la proximité de la coloration émotionnelle de l'amorce et de la cible). Si ces hypothèses étaient vérifiées, la prise en compte du contexte émotionnel durant ces deux phases — encodage et récupération — pourrait fournir des indices pertinents pour la prédiction de la mémorabilité des images.

Bilan

Dans la suite de ce chapitre, nous proposons plusieurs méthodes pour évaluer l'influence du contexte (global et local) de présentation des images sur leur mémorabilité, basées sur : (1) la fréquence d'occurrence dans une tâche de reconnaissance du type de scène représentée par une image, (2) le degré de cohérence du contexte émotionnel local de récupération (i.e. de reconnaissance) de l'image, et (3) la similarité des contextes émotionnels d'encodage et de récupération de l'image. Nous utilisons les données recueillies dans l'expérience décrite dans le chapitre 5, dans laquelle nos participants ont réalisé une tâche de reconnaissance et ont évalué les images sur les dimensions d'arousal et de valence. Nous attendons de l'analyse de ces données qu'elle révèle éventuellement des effets de contexte sur la mémorabilité des images, que nous pourrions prendre en

compte pour ajuster les scores de mémorabilité moyens générés par notre modèle à apprentissage profond, MemoNet.

9.2 Fréquence d'une scène sur la mémorabilité d'une image

Le contexte de présentation des images peut influencer leur mémorisation et leur récupération. Dans les deux tests de mémoire utilisés pour obtenir des scores de mémorabilité pour notre base de données (voir chapitre 5), l'ordre de présentation des images était aléatoire pour chaque participant. L'objectif était de nous assurer que les performances de mémoire mesurées ne dépendraient pas de l'ordre dans lequel les images étaient présentées. De plus, lors d'un test de mémoire, un participant ne voyait pas plus de trois images d'une même catégorie (p. ex. trois serpents). En effet, on peut aisément imaginer que la rareté de l'image la rende plus mémorable ; pour prendre un exemple extrême, une image de serpent présentée parmi 100 images de voitures aura certainement plus de chance d'être reconnue que si elle est présentée parmi 100 images de serpent. De plus, la multiplication des images d'une même catégorie pourrait conduire les participants à reconnaître à tort une image jamais vue de cette catégorie, à cause de sa similarité en matière de contenu avec les autres images de cette catégorie (on notera qu'il commettra alors une fausse alarme). Toutefois, malgré cette précaution, il est possible que les effets causés par la fréquence d'apparition de la catégorie d'une image sur sa mémorabilité apparaisse même lorsque cette fréquence ne dépasse pas trois occurrences par test de mémoire.

Pour tester cette hypothèse, nous avons calculé la fréquence d'apparition moyenne de chaque catégorie d'images (en utilisant les catégories relatives au contenu sémantique fournies par (Lang et al., 2008), soit 339 catégories différentes comprenant les 600 images utilisées dans notre étude) pour les différents jeux d'images vus par les participants. Une catégorie d'images pouvait apparaître une, deux, ou trois fois dans un même test de mémoire.

Pour la première tâche de reconnaissance, les moyennes de mémorabilité des images étaient les suivantes : $\mu = 0.74$ pour les images dont la catégorie d'appartenance n'apparaissait qu'une fois dans le test, $\mu = 0.77$ pour les images dont la catégorie d'appartenance apparaissait deux fois dans le test, et $\mu = 0.84$ pour les images dont la catégorie d'appartenance apparaissait trois fois dans le test. Nous avons réalisé un test de Kruskal-Wallis pour comparer les médianes des scores de mémorabilité des trois groupes de fréquence d'apparition par test de mémoire (avec trois modalités : 1, 2 ou 3 apparitions de la catégorie dans le test). Le résultat du test ($\chi^2(2) = 6.19, p < .05$) révèle un effet principal significatif de la fréquence d'apparition de la catégorie d'images sur la mémorabilité. Nous avons effectué des comparaisons multiples, en utilisant la mé-

thode de Tukey, pour tester chaque groupe contre les deux autres : les résultats révèlent une différence significative entre les groupes 1 et 3. Cela signifie que les images dont la fréquence d'apparition de la catégorie à laquelle elles appartiennent est la plus élevée tendent à avoir une mémorabilité significativement plus élevée que les images dont la fréquence d'apparition de la catégorie à laquelle elles appartiennent est la plus faible.

Nous avons observé le même pattern pour les scores de mémorabilité calculés à partir des résultats à la seconde tâche de reconnaissance ($\chi^2(2) = 7.67, p < .05$). Les moyennes de mémorabilité des images associées à chacune des conditions étaient les suivantes : $\mu = 0.5359$ pour les images dont la catégorie d'appartenance n'apparaissait qu'une fois dans le test, $\mu = 0.5428$ pour les images dont la catégorie d'appartenance apparaissait deux fois dans le test, et $\mu = 0.6711$ pour les images dont la catégorie d'appartenance apparaissait trois fois dans le test.

En somme, les résultats suggèrent que plus une catégorie d'image apparaît fréquemment, et ce même lorsqu'il ne s'agit au maximum que de trois apparitions d'une même catégorie pour 200 images composant une tâche de reconnaissance, plus la mémorabilité des images appartenant à cette catégorie tend à augmenter. Cependant, cette interprétation pourrait être trompeuse, pour au moins deux raisons.

D'une part, la différence de mémorabilité moyenne des images observée entre les trois conditions de fréquence d'apparition des catégories d'images dans un test pourrait être due au fait que les participants avaient plus tendance à reconnaître une image dont la catégorie était sur-représentée, indépendamment de la mémorabilité intrinsèque de l'image. Autrement dit, un participant aura peut-être plus eu tendance à répondre « déjà vue » pour des images répétées et non répétées dont la catégorie d'appartenance a déjà été vue auparavant, que pour des images dont la catégorie est vue pour la première fois. Si cette hypothèse est vraie, cela devrait se traduire par une augmentation des fausses alarmes (FA) pour les catégories les plus représentées. Pour éprouver le bien-fondé de cette hypothèse, nous avons testé les taux de FA associés aux images de remplissage en fonction de la fréquence d'apparition des catégories d'appartenance (de la même manière dont nous avons testé les scores de mémorabilité des images cibles). Pour le premier test de mémoire, les pourcentages moyens de FA pour les images des trois conditions étaient les suivants : 0.44% pour les images dont la catégorie d'appartenance n'apparaissait qu'une fois dans le test, 2% pour les images dont la catégorie d'appartenance apparaissait deux fois dans le test, et 4.3% pour les images dont la catégorie d'appartenance apparaissait trois fois dans le test. Un test de Kruskal-Wallis révèle un effet principal de la fréquence de la catégorie d'image sur le taux de FA ($\chi^2(2) = 59.18, p < .001$). Des comparaisons multiples révèlent des différences significatives entre chacun des groupes. Nous avons trouvé un pattern similaire pour le second test de mémoire ($\chi^2(2) = 28.90, p < .001$), avec des comparaisons multiples révélant une différence significative entre le groupe d'images dont la catégorie était la moins fréquente et les deux autre groupes. Les taux moyens de FA pour chaque

condition étaient les suivants : 1.86% pour la première condition, 4.82% pour la seconde condition, et 6.55% pour la troisième condition (i.e. pour laquelle la catégorie de chaque image apparaissait trois fois dans le test de mémoire). Ces résultats signifient que la sur-représentation d'une catégorie d'images, dans notre étude, s'accompagnait d'une probabilité plus élevée de commettre une FA sur une image appartenant à cette catégorie. Il faut noter que le taux moyen de FA était bas dans notre étude : 0.89% pour le premier test de reconnaissance, et 3.69% pour le second test, un jour plus tard. Cela suggère que les variations significatives de la mémorabilité des images en fonction de la fréquence d'apparition de la catégorie d'appartenance des images dans un test de mémoire sont (en partie) dues au fait que les participants avaient plus tendance à reconnaître les images dont les catégories étaient les plus représentées, ce qui augmentait leur chance de détecter une image à raison comme à tort.

D'autre part, l'émotion véhiculée par une image étant étroitement liée à sa mémorabilité, les différences de mémorabilité moyenne des images observées selon que la fréquence d'apparition de la catégorie d'appartenance des images est plus ou moins élevée pourraient être dues à un effet confondu de l'émotion suscitée par les images, si la répartition des images dans l'espace émotionnel n'est pas la même pour les différentes conditions de fréquence des catégories d'appartenance. L'hypothèse d'une surreprésentation des images véhiculant des émotions intenses (i.e. négatives ou positives, suscitant un arousal fort) dans les catégories d'images les plus fréquentes de l'IAPS, et par conséquent du sous-ensemble d'images que nous avons utilisé dans notre étude, n'est en effet pas inconcevable dans une base créée, à l'origine, expressément pour l'induction d'émotion ; or ce type d'image est aussi le plus mémorable. Pour tester cette hypothèse, nous nous sommes intéressés aux scores d'arousal et de valence des 150 images évaluées par nos participants, en fonction de la fréquence d'apparition de chaque catégorie d'images dans un test de mémoire. Les scores moyens d'arousal pour chaque condition étaient les suivants : $\mu = 3.54$ pour les images dont la catégorie d'appartenance n'apparaissait qu'une fois par test de mémoire, $\mu = 4.11$ pour les images dont la catégorie d'appartenance apparaissait deux fois par test de mémoire, et $\mu = 4.26$ pour les images dont la catégorie d'appartenance apparaissait trois fois par test de mémoire. Un test de Kruskal-Wallis a révélé un effet global de la fréquence de la catégorie de l'image dans les jeux d'images utilisés sur le score d'arousal d'une image ($\chi^2(2) = 6.98, p < .05$). Les scores moyens de valence pour chaque condition étaient les suivants : $\mu = 4.47$ pour la première condition, $\mu = 4.39$ pour la seconde condition, et $\mu = 4.46$ pour la troisième condition. Un test de Kruskal-Wallis n'a révélé aucun effet principal de la fréquence de la catégorie d'appartenance de l'image sur son score de valence ($\chi^2(2) = 0.05, p = .9748$). Ces résultats signifient que les images des catégories qui apparaissaient plus fréquemment ont suscité en moyenne un arousal plus fort ; et elles tendaient également, comme on l'a vu, à être plus mémorables. Or, nous avons montré dans le chapitre 7 une corrélation linéaire positive entre l'arousal et

la mémorabilité (pour rappel : $r = .23, p < .01$ pour le premier test de mémoire, et $r = .42, p < .001$ pour le deuxième test). Il s'ensuit que l'arousal (plutôt que la fréquence d'apparition de la catégorie d'appartenance de l'image dans un test de mémoire) pourrait être la cause d'une mémorabilité plus élevée pour les images dont la catégorie apparaissait la plus fréquemment.

On peut noter, à ce sujet, que lorsque nos participants ont évalué le degré d'arousal associé aux images, les catégories d'appartenance de ces images avaient déjà été vues dans les tests de mémoire, un nombre de fois plus ou moins grand (i.e. une à trois fois pour chacun des deux tests de mémoire qui précédaient la tâche de reconnaissance). Or, il est possible que la répétition du visionnage d'une même catégorie d'images tende à baisser l'intensité de l'émotion induite chez l'observateur par les images de cette catégorie (p. ex. le cinquième accident de voiture pourrait susciter une émotion moins intense que le premier). Dans cette hypothèse, les participants auront évalué les images des catégories les plus représentées comme causant un degré d'arousal moindre que s'ils avaient évalué ces mêmes images alors que leur catégorie d'appartenance eût été vue pour la première fois. Il faut noter que cet effet, s'il était avéré, agissant à l'opposé du résultat suivant lequel les scores d'arousal sont plus élevés dans les catégories les plus fréquentes, renforce encore l'idée d'une sur-représentation des images suscitant un degré d'arousal élevé dans les catégories d'images les plus fréquentes dans la collection d'images que nous avons utilisée dans notre étude, et s'ajoute d'autant à l'argument selon lequel l'arousal plutôt que la fréquence d'apparition des catégories d'images a causé une augmentation de la mémorabilité des images dont les catégories sont les plus représentées.

9.3 Effets contextuels de l'émotion sur la récupération mnésique

Le contexte émotionnel dans lequel est présenté une image est, comme nous l'avons précédemment évoqué, susceptible d'influencer sa mémorabilité. Dans cette section, nous nous intéressons à deux types d'effet contextuel potentiels de l'émotion sur la mémoire : la proximité en matière d'émotion véhiculée entre une image cible et l'image qui la précède sur la probabilité que l'image cible soit reconnue, et la récupération mnésique dépendante de la similarité des contextes émotionnels d'encodage et de récupération. Nous utilisons, dans les analyses effectuées dans cette section, les scores d'arousal et de valence fournis par (Lang et al., 2008). La raison en est que nos analyses nécessitent de prendre en compte l'ensemble des images vues dans les tests de mémoire, et non seulement les images cibles, qui seules ont été évaluées par nos participants sur les dimensions d'arousal et de valence. Il est bon de rappeler ici que la corrélation entre les scores d'émotion obtenus dans notre expérience décrite au chapitre 5 et ceux obtenus

par Lang *et al.* est de .95 pour la valence, et de .69 pour l'arousal.

9.3.1 Effets contextuels de l'émotion lors de la récupération mnésique

Pour étudier les effets contextuels de l'émotion sur la probabilité de reconnaître une image répétée, nous avons, dans un premier temps, décidé d'évaluer la similarité en matière d'arousal et de valence entre chaque image répétée et l'image qui la précédait, dans la première tâche de reconnaissance.

Pour chaque image cible *i*, nous avons d'abord calculé l'écart entre son score d'arousal et le score d'arousal de l'image qui précédait sa répétition, tel que : $\Delta_{Ar_{R_i}} = Ar_i - Ar_{i-1}$, avec Ar le score d'arousal de l'image considérée, et R pour indiquer que nous nous intéressons au contexte émotionnel de récupération de l'image, c'est-à-dire au moment de la répétition de l'image cible et non au moment de sa première occurrence, qui a pour objectif son encodage par le participant (voir la figure 9.1, encadré vert). Nous avons procédé de même pour la valence, en calculant pour chaque image cible un Δ_{Val_R}.

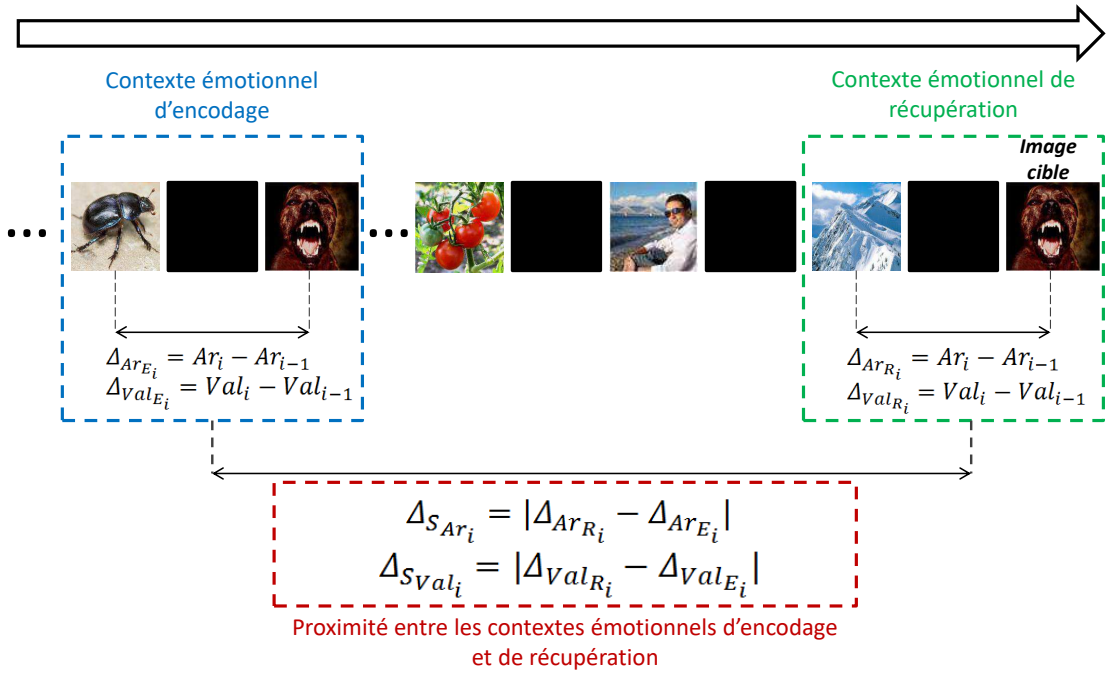


FIGURE 9.1 – Prise en compte des contextes émotionnels d'encodage (encadré bleu) et de récupération (encadré vert) des images. Exemple pour une image i.

Ensuite, nous avons calculé pour chaque participant, pour chaque image cible, un $\Delta_{Mémorabilité}$, tel que $\Delta_M = D_{i,j} - \mu_i$, avec $D_{i,j}$ la détection correcte (1) ou l'oubli

incorrect (0) de l'image répétée *i* par le sujet *j*, et μ_i le score moyen de mémorabilité de l'image (i.e. le taux moyen de détection correcte sur l'ensemble des répétitions de l'image). L'intérêt de retrancher la mémorabilité moyenne de l'image aux valeurs de détection et d'oubli réside dans le fait que les Δ_{Ar} ne sont pas indépendants des valeurs d'arousal des images ($r = .71, p < .001$), ni les Δ_{Val} des valeurs de valence des images ($r = -.19, p < .001$). Comme la mémorabilité n'est pas elle-même indépendante de l'arousal et de la valence, il nous a fallu mettre en place cette soustraction afin de sortir de l'analyse la part des $\Delta_{Ar,Val}$ liée aux variations d'arousal et de valence des images.

Nous n'avons pas observé de corrélations significatives entre les Δ_{Ar_R} et les Δ_M ($r(\Delta_{Ar}, \Delta_M) = -.03, p = .21$), ni entre les Δ_{Val_R} et les Δ_M ($r(\Delta_{Val}, \Delta_M) = .00, p = .97$). Nous avons réalisé une analyse similaire à partir des résultats à la seconde tâche de reconnaissance, et obtenu les coefficients de corrélation suivants : $r(\Delta_{Ar}, \Delta_M) = -.01, p = .52$, et $r(\Delta_{Val}, \Delta_M) = .00, p = .96$. L'absence de résultat ne permet évidemment pas de conclure à une absence d'effet du contexte émotionnel de présentation d'une image sur la probabilité qu'elle soit reconnue. Nous revenons plus en détail sur ce point dans la section consacrée à la discussion, et proposons d'autres manières de prendre en compte le contexte émotionnel lors de la récupération d'une image.

9.3.2 Récupération mnésique dépendante de la similarité des contextes émotionnels d'encodage et de récupération

Suivant l'effet de récupération dépendante du contexte d'encodage, plus le contexte émotionnel d'encodage d'une image est proche du contexte émotionnel de récupération, plus l'image aura de chance d'être reconnue.

Pour obtenir une mesure du degré de similarité en matière d'arousal et de valence entre le contexte émotionnel d'encodage et le contexte émotionnel de récupération, nous avons calculé pour chaque image cible i l'écart entre son score d'arousal et le score d'arousal de l'image qui précédait sa première occurrence (i.e. son encodage) sur le modèle des Δ à la récupération, tel que $\Delta_{Ar_{E_i}} = Ar_i - Ar_{i-1}$, avec Ar le score d'arousal de l'image considérée, et E pour indiquer que nous nous intéressons au contexte émotionnel d'encodage de l'image (voir la figure 9.1, encadré bleu). Nous avons procédé de la même façon pour obtenir un Δ_{Val_E} pour chaque image cible.

Ensuite, nous avons calculé pour chaque image cible le degré de proximité des contextes d'encodage et de récupération en matière d'arousal, tel que : $\Delta_{S_{Ar_i}} = \mid \Delta_{Ar_{R_i}} - \Delta_{Ar_{E_i}} \mid$, avec S pour indiquer la similarité entre les contextes d'encodage et de récupération (voir la figure 9.1, encadré rouge). Nous avons procédé de la même façon pour obtenir un Δ_{Val_S} pour chaque image cible.

Pour le premier test de mémoire, nous n'avons pas trouvé de corrélation significative entre la performance de reconnaissance d'un participant par rapport à la performance de reconnaissance moyenne et la proximité des contextes d'encodage et de récupération,

que ce soit en matière d'arousal ou de valence, ni pour le premier test de mémoire ($r(\Delta_{S_{Ar}}), \Delta_M) = -.003, p = .88$; $r(\Delta_{S_{Val}}), \Delta_M) = -.001, p = .9578$), ni pour le second test ($r(\Delta_{S_{Ar}}), \Delta_M) = .004, p = .8426$; $r(\Delta_{S_{Val}}), \Delta_M) = .008, p = .6914$). L'absence de résultat ne permet pas de conclure à une absence d'effet; dans la partie consacrée, nous discutons d'autres méthodes qui pourraient permettre de révéler d'éventuels effets que notre méthode n'aura pas permis de mettre en lumière.

9.4 Discussion et résultats complémentaires

Dans les sections précédentes de ce chapitre, nous avons cherché à mettre en lumière des effets du contexte de présentation des images sur leur mémorabilité, dans l'objectif de les prendre en compte pour prédire la mémorabilité des images. Les résultats ne sont pas concluants. Dans la discussion qui suit, nous cherchons à expliquer pourquoi, et apportons des résultats complémentaires pour étayer nos propos.

Effet du contexte global sur la mémorabilité des images

Isola *et al.*, qui ont trouvé une corrélation négative (faible) entre la fréquence du type de scène représentée dans les images et la mémorabilité moyenne des images, ont recommandé de mesurer la mémorabilité sur de nouveaux jeux d'images, pour évaluer à quel point leurs résultats sont généralisables à d'autres jeux de données (Isola et al., 2011b). Nos résultats, quoiqu'ils aient paru au début de nos analyses aller en sens inverse de ceux de Isola *et al.*, ne nous ont pas permis de conclure dans ce sens. En effet, les effets confondus de l'émotion véhiculée par les images et du positionnement des participants en matière de risque pris (qui se traduisait par un taux de FA particulier) ne nous ont pas permis de faire une telle interprétation. Nous avons montré dans le chapitre précédent qu'il était raisonnable, dans une base de données destinée à l'étude de la mémorabilité des images, de s'intéresser à la répartition des images dans l'espace arousal-valence. C'est d'autant plus vrai que les résultats présentés dans ce chapitre suggèrent que la fréquence d'occurrence d'un type de scène dans un jeu d'images et les émotions véhiculées par les images peuvent interagir, et que cette interaction influence la mémorabilité des images. Par exemple, dans l'IAPS, les catégories d'images ne sont pas également réparties dans l'espace arousal-valence, ce qui est logique puisque l'émotion suscitée par les images a été évaluée après avoir constitué le jeu d'images, qui n'a pas été créé pour obtenir une telle égalité. Si cela est vrai dans l'IAPS, ce pourrait également être vrai dans la base de (Isola et al., 2011b).

Effet contextuel de l'émotion sur la mémoire lors de la récupération

Nous avons évalué l'influence du contexte émotionnel de présentation des images sur leur mémorabilité à partir de deux approches différentes. La première approche visait à mettre en lumière un éventuel effet de congruence émotionnelle dans une tâche de reconnaissance d'images. À travers la seconde approche, nous avons cherché à déterminer si la similarité des contextes émotionnels d'encodage et de récupération d'une image influençait sa mémorabilité. Nos analyses n'ont pas abouti à des résultats significatifs. Cela ne signifie pas que ces effets n'existent pas : nous n'avons peut-être simplement pas réussi à les montrer. D'autant que notre expérience n'a pas été spécifiquement conçue pour étudier ces effets. En effet, les conditions dans lesquelles nous avons cherché à montrer de tels effets sont très différentes des conditions dans lesquelles ils ont été précédemment montrés. En particulier, nous ne nous sommes pas intéressés à l'humeur des participants, mais aux émotions induites chez eux par les images d'une tâche de reconnaissance. Invoquer ces effets nous aura permis, simplement, de commencer quelque part l'étude des effets contextuels de l'émotion sur la mémorabilité des images, avec l'idée de les prendre en compte dans les modèles computationnels.

Typiquement, l'effet de congruence émotionnelle se traduit par une tendance chez un individu à récupérer des souvenirs dont la coloration émotionnelle coïncide avec la tonalité affective de son état émotionnel persistant. Nous avons fait l'hypothèse que l'effet de congruence émotionnelle pouvait également apparaître lorsque, dans une tâche de reconnaissance d'images, il y a congruence entre une image cible et l'image qui la précède. D'une part, du rappel de souvenirs autobiographiques, nous sommes passés à la reconnaissance d'images : donc, à la fois le matériel et le type de tâche différaient des études qui ont précédemment montré cet effet. D'autre part, d'une congruence avec une humeur (i.e. un état émotionnel durable ; p. ex. la dépression, l'anxiété), nous somme passés à une congruence avec une émotion passagère (induite par la présentation d'une image).

Pour ce qui est du premier point, on peut imaginer que, dans une tâche de reconnaissance, les individus sont moins susceptibles dans leur recherche en mémoire d'être influencés par leur état émotionnel que dans une tâche de rappel. En effet, en reconnaissance, l'accès au souvenir à récupérer est facilité par la présentation du stimulus et, en particulier, l'émotion qu'il suscite, alors qu'en rappel aucun indice n'est donné sur la coloration émotionnelle du souvenir à récupérer. Dans une telle hypothèse, l'effet de congruence émotionnelle devrait être plus susceptible d'apparaître en tâche de rappel qu'en tâche de reconnaissance. Pour ce qui est des images, elles pourraient fournir un certains nombre d'indices visuels, dont on peut imaginer, en raison de l'importance de la vision chez l'homme (Chalupa and Werner, 2004), qu'ils sont utilisés en priorité par les systèmes mnésiques pour accéder aux souvenirs, qui laissent alors moins de place à l'émotion pour influencer la recherche en mémoire que lorsque les individus sont libres de chercher les souvenirs qui les intéressent.

Pour ce qui est du second point, l'état émotionnel dont nous parlons dans notre étude est celui, transitoire, induit par une image que le participant visionne, et non pas un état émotionnel persistant tel que l'humeur. Or, il n'est pas certain que l'état émotionnel induit par une image dure suffisamment longtemps pour entrer en interaction avec l'état émotionnel induit par l'image suivante (i.e. la cible), lorsque, comme dans notre étude, les images sont présentées deux secondes chacune, avec un intervalle inter-stimuli d'une seconde. Peut-être aurait-il d'ailleurs été sage de parler d'effet d'amorçage affectif sur la mémoire. L'amorçage désigne un effet de mémoire implicite, selon lequel la présentation d'un stimulus (l'amorce) influence le traitement d'un stimulus présenté ensuite (la cible) ; l'amorçage affectif désigne un amorçage déterminé par la charge émotionnelle des stimuli. Par exemple, Avero et Calvo ont montré un effet d'amorçage affectif pour des images de l'IAPS : lorsque l'amorce (une image soit positive, soit négative) était congruente en matière de valence avec l'image cible, le temps de réponse pour décider si la cible était positive ou négative était plus court que lorsque la cible n'était pas congruente avec l'amorce (Avero and Calvo, 2006). Ce résultat a été montré de nombreuses fois et est considéré comme robuste (p. ex. (Fazio et al., 1986, Bargh et al., 1992, Hermans et al., 1994)). Sur la base de ces travaux, nous aurions pu émettre l'hypothèse suivante : lorsqu'il y a congruence émotionnelle entre l'image cible et l'image qui la précède (qui sert d'amorce), la récupération de la cible est plus aisée ; celle-ci sera donc plus probablement reconnue à cause de la contrainte de temps dans notre étude (pour rappel, un participant avait deux secondes pour reconnaître une image dans notre tâche). Dans les études portant sur l'effet d'amorçage affectif (que le matériel soit des mots ou des images), le SOA (ou *Stimulus Onset Asynchrony*, qui désigne le temps qui s'écoule entre le début de la présentation d'un premier stimulus et le début de la présentation d'un second stimulus, est typiquement de 300 ms (voir à ce sujet la vue d'ensemble proposée par (Hermans et al., 2001)). Certains auteurs ont cependant fait varier ce SOA et observé que l'effet d'amorçage affectif observé pour un SOA de 300 ms n'apparaissait plus pour un SOA de 1000 ms (Fazio et al., 1986, Eelen, 1998, Hermans et al., 1994). Or, dans notre étude, le SOA était de 3000 ms.

En rapport avec ce qui vient d'être dit, il est intéressant de faire deux analyses complémentaires. Une première analyse portant sur les temps de réponse des participants pour reconnaître une image répétée, en proposant l'hypothèse qu'ils sont plus courts lorsqu'il y a congruence émotionnelle entre l'amorce et la cible. L'effet d'amorçage affectif a, en effet, été étudié sur le temps de traitement des stimuli cibles dans une tâche de catégorisation (p. ex. le participant doit dire si le stimulus cible est positif ou négatif), et nous avons fait l'hypothèse qu'il pourrait se traduire par une augmentation de la probabilité de reconnaître une image dans une tâche de reconnaissance. Une seconde analyse portant sur l'effet de congruence émotionnelle entre l'image cible et l'humeur des participants au moment de la tâche, telle que nous l'avons mesurée au début de chaque tâche de reconnaissance à l'aide de la matrice de l'humeur, ainsi qu'entre l'image cible

et l'humeur durable, ou *trait*, des participants, telle que nous l'avons mesurée à l'aide de l'I-PANAS-SF (qui porte sur l'année écoulée). Dans ces conditions — particulièrement dans l'étude d'une congruence cible-I-PANAS-SF —, nous serions plus proches des conditions dans lesquelles l'effet de congruence émotionnelle a été montré (pour des souvenirs autobiographiques chez des individus durablement dépressif, anxieux, heureux, etc.). Les résultats de ces analyses sont exposés dans les prochains paragraphes.

L'analyse des temps de réponse des participants a porté sur les images répétées qui ont été reconnues. Nous avons calculé un coefficient de corrélation entre, d'une part, les valeurs absolues des Δ_{Ar_R} et les Δ_{Val_R} (calculées dans la section précédente), qui correspondent au degré de congruence entre l'amorce (i.e. l'image qui précède l'image cible) et la cible en matière d'arousal et de valence (plus le Δ est faible, plus la congruence est faible), et, d'autre part, le temps Tps mis pour répondre par chaque participant sur chaque image (à partir du début de la présentation de l'image cible ; ce temps maximum est de trois secondes, comme nous prenions en compte la réponse de nos participants sur l'écran noir suivant l'image cible). Pour le premier test de mémoire (voir figure 9.2, partie haute), nous n'avons pas observé de corrélation significative entre les $\mid \Delta_{Ar_R} \mid$ et les temps de réponse ($r(\Delta_{Ar_R}, Tps) = -.000, p = .994$), ni entre les $\mid \Delta_{Val_R} \mid$ et les temps de réponse ($r(\Delta_{Ar_R}, Tps) = -.013, p = .565$). Pour le second test de mémoire (voir figure 9.2, partie basse), nous n'avons pas observé non plus de corrélation significative entre les $\mid \Delta_{Ar_R} \mid$ et les temps de réponse ($r(\Delta_{Ar_R}, Tps) = -.02, p = .465$), ni entre les $\mid \Delta_{Val_R} \mid$ et les temps de réponse ($r(\Delta_{Ar_R}, Tps) = -.045, p = .097$). Cette absence de corrélation significative n'est pas étonnante puisque, comme nous l'avons dit, l'effet d'amorçage affectif disparaît généralement lorsque le SOA est égal à seulement 1000 ms ; or il était de 3000 ms dans notre étude.

Puisque nous parlons de temps d'affichage des images cibles, il est intéressant de rappeler que les scores de mémorabilité de l'ensemble des images utilisées pour la prédiction de la mémorabilité des images (ceux collectés dans notre étude compris) ont été calculés à partir de mesures de performances de mémoire à temps contraint. Or, la contrainte de temps a pu influencer la mémorabilité des images. Imaginons, par exemple, que les participants aient eu deux fois plus de temps pour répondre, ou même une infinité de temps : on peut penser qu'en ayant du temps supplémentaire pour sonder leur mémoire, ils auraient reconnu les images répétées plus souvent. Il serait intéressant qu'une prochaine étude visant à créer un nouveau jeu de données destiné à l'étude de la mémorabilité des images teste cette hypothèse, en imposant aux participants d'autres contraintes temporelles, ou pas du tout.

Pour étudier l'effet de congruence émotionnelle entre l'image cible et l'état émotionnel du participant, soit mesuré par la matrice de l'humeur, soit par l'I-PANAS-SF, sur la probabilité de reconnaître une image répétée, nous avons ajouté puis testé plusieurs hypothèses a posteriori. En conséquence, nous avons corrigé le seuil de rejet des

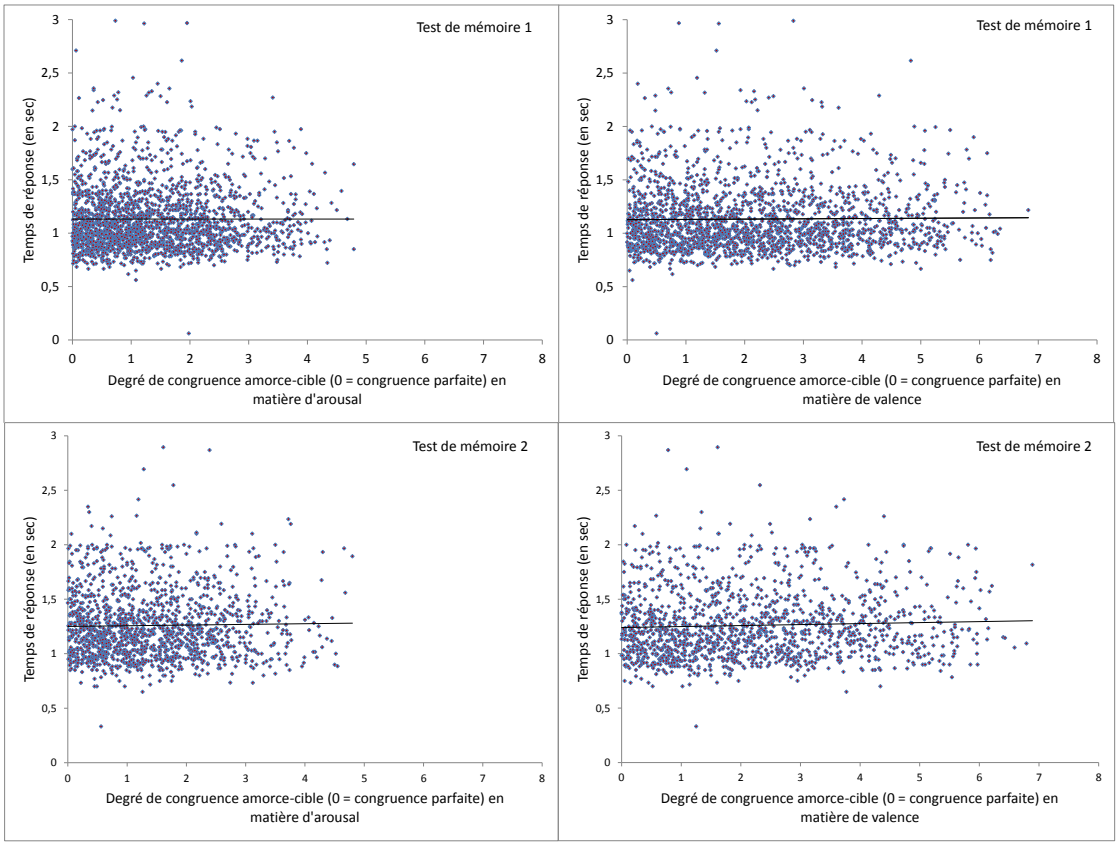


FIGURE 9.2 – Graphiques des temps de réponses des participants pour les images qu'ils ont reconnues en fonction du degré de congruence émotionnelle entre l'image considérée (la cible) et l'image qui la précède (l'amorce). Chaque point représente une image répétée reconnue par un participant. Le degré de congruence émotionnelle correspond aux valeurs absolues des Δ_{Ar_R} et Δ_{Val_R} précédemment calculées ; il quantifie la proximité en matière d'arousal et de valence entre l'amorce et la cible. Le degré de congruence émotionnelle théorique est de 8, puisqu'il ne peut pas y avoir entre deux images dont la charge émotionnelle a été évaluée par les échelles SAM (Bradley and Lang, 1994) une distance supérieure à celle-ci (un tel degré signifierait que l'une des deux images — l'amorce ou la cible — a une valeur sur la dimension considérée de 1, et l'autre de 9). On remarquera que les scores d'arousal étant globalement de valeur moins extrêmes (sur les échelles SAM) que les scores de valence, le degré de congruence tel que nous l'avons calculé tend en conséquence à être moins fort pour l'arousal que pour la valence. Les données du premier test de mémoire (en haut), réalisé quelques minutes après l'encodage mnésique, et du second test de mémoire (en bas), réalisé un jour plus tard, sont utilisées. On remarquera que, le taux de reconnaissance moyen étant plus faible au second test qu'au premier, le nombre de données disponibles pour mesurer les temps de réponse est plus faible pour le second test que pour le premier (le temps de réponse moyen tend également à être un peu plus long au second test). Le temps de réponse est de $3s$ maximum (les $2s$ d'affichage de l'image cible, plus les $1s$ d'écran noir suivant l'image, sur lesquelles les réponses étaient encore enregistrées). Les lignes droites continues représentent les courbes de tendance linéaire des données.

hypothèses nulles en appliquant une correction de Bonferroni, de telle sorte qu'on nous rejetons dans la suite de nos analyses H_i si $p_i <= \alpha/m$, avec H_i une hypothèse nulle ajoutée a posteriori, $p-i$ la valeur-p correspondant à l'hypothèse considérée, α le risque de première espèce a priori (i.e. le risque de rejeter l'hypothèse nulle alors qu'elle est vraie), et m le nombre d'hypothèses ajoutées a posteriori. Nous testons ci-après six hypothèses ajoutées a posteriori, de sorte qu'une valeur-p devra être inférieure ou égale à $0.05/6$, soit 0.0083, pour décider qu'on rejette l'hypothèse nulle au profit de l'hypothèse alternative. Le sujet 44 a été retiré des analyses, comme il n'a pas passé la tâche d'évaluation de l'arousal et de la valence suscités par les images. Les analyses portent donc sur 49 participants, chacun d'entre eux ayant eu à reconnaître 100 images cibles (50 lors du premier test de mémoire, et 50 lors du second), puis à évaluer ces 100 images sur les dimensions d'arousal et de valence.

Pour tester l'effet de la congruence émotionnelle entre une image cible et l'état émotionnel du participant mesuré par la matrice de l'humeur sur la probabilité de reconnaître un image, nous avons d'abord calculé, pour chaque image cible, pour chaque participant, un $\Delta_{Ar_{MH}}$ représentant la distance en matière d'arousal qui sépare la notation de l'image réalisée par le participant sur cette dimension et la valeur d'arousal que ce participant s'est auto-attribuée à l'aide de la matrice de l'humeur, tel que : $\Delta_{Ar_{MH_i}} =\mid (NoteImage - NoteMatrice) \mid$. Nous avons procédé de la même manière pour obtenir les $\Delta_{Val_{MH}}$. Il faut rappeler ici que les mesures d'arousal et de valence réalisées à l'aide des échelles SAM et de la matrice de l'humeur sont similaires : le participant attribue dans les deux cas une note de 1 à 9 pour chacune des deux dimensions, et les notes ont une signification similaire. Nous avons fait ce calcul pour les deux tests mémoire ; les participants ayant évalué leur état émotionnel sur la matrice de l'humeur au début de chacun des deux tests de mémoire, les résultats de la matrice pris en compte dans l'analyse sont ceux qui correspondent au test de mémoire considéré. On remarquera également que, contrairement aux analyses présentées dans la section précédente de ce chapitre, qui se basaient sur les scores d'émotion des images fournis par (Lang et al., 2008) en raison du fait que nos participants n'ont pas évalué les images de remplissage (dont les scores d'émotion étaient nécessaires dans ces analyses), la présente analyse se base sur les notes attribuées par chaque participant à chacune des images cibles qu'il a vues. Un $\Delta_{Val_{MH}}$ (ou un $\Delta_{Ar_{MH}}$) indique le degré de congruence entre l'image cible et l'état émotionnel du participant au moment de la tâche de reconnaissance : étant donné l'effet de congruence émotionnelle, notre hypothèse opérationnelle est que ces deltas devraient être plus faibles pour les images reconnues que pour les images oubliées.

Les résultats sont présentés dans la partie gauche de la figure 9.3 ; pour la lisibilité de la figure, l'axe des ordonnées correspond au degré de congruence émotionnelle, entre 0 et 1, tel que $C_i = (8 - \Delta_{Ar_{MH_i}})/8$, avec C_i la congruence pour le $\Delta_{Ar_{MH_i}}$ (la valeur 8 étant la valeur maximum que peut prendre un $\Delta_{Ar_{MH}}$). Un test U de Mann-

Whitney n'indique aucune différence significative entre les $\Delta_{Ar_{MH}}$ pour les images répétées reconnues (moyenne $\mu = 2.993$) et les $\Delta_{Ar_{MH}}$ pour les images répétées oubliées ($\mu = 3.034$) pour le premier test de mémoire ($Z = 0.652, p = .4151$). En revanche, pour le second test de mémoire, un test similaire indique que les $\Delta_{Ar_{MH}}$ sont significativement plus faibles pour les images répétées reconnues ($\mu = 2.982$) que pour les images répétées oubliées ($\mu = 3.276$) ($Z = 3,77, p(= .00016) < .0083$). Pour la valence, un test U de Mann-Whitney n'indique aucune différence significative entre les $\Delta_{Val_{MH}}$ pour les images répétées reconnues ($\mu = 2.757$) et les $\Delta_{Val_{MH}}$ pour les images répétées oubliées ($\mu = 2.497$) pour le premier test de mémoire ($Z = -2.528, p = .01147$); une tendance semble cependant se dessiner. Pour le second test de mémoire, un test statistique similaire indique que les $\Delta_{Val_{MH}}$ sont significativement plus élevés pour les images répétées reconnues ($\mu = 2.765$) que pour les images répétées oubliées ($\mu = 2.456$) ($Z = -3,334, p(= .00086) < .0083$).

Si nous n'avons pas obtenu de résultats significatifs pour le premier test de mémoire, pour le second nos résultats vont dans le sens d'un effet de congruence émotionnelle, qui serait facilitateur de la récupération mnésique pour l'arousal, et contraire pour la valence. Ce dernier résultat, qui va à l'inverse de notre hypothèse, suggère que le contraste entre la coloration émotionnelle d'une image et la tonalité affective de l'état émotionnel d'un individu améliore la récupération des images dans le cadre d'un test de reconnaissance. Les résultats suggèrent également, puisque les effets paraissent au second test de mémoire et pas au premier test (où il y a cependant une tendance pour la valence), que les effets contextuels de l'émotion sur la récupération mnésique sont d'autant plus forts que la durée de la rétention mnésique est élevée. Dans tous les cas, il s'agit ici d'un travail exploratoire, et nos résultats appellent une corroboration par de nouvelles études.

Pour tester l'effet de la congruence émotionnelle entre une image cible et l'état émotionnel du participant mesuré par l'I-PANAS-SF sur la probabilité de reconnaître une image, nous ne nous sommes intéressés qu'à la dimension émotionnelle de valence. En effet, contrairement à la matrice de l'humeur, les résultats donnés par l'I-PANAS-SF ne sont pas du même format que ceux donnés par les échelles SAM. L'I-PANAS-SF donne à chaque individu un score sur deux dimensions, « positive » et « négative ». Pour l'analyse qui suit, nous avons créé deux catégories : la première catégorie regroupe les participants qui ont un score supérieur sur la dimension « positive » que sur la dimension « négative » ; la seconde catégorie regroupe les participants qui montrent un pattern inverse (cette dernière catégorie compte seulement 5 participants sur les 49, et 500 observations — i.e. 500 images répétées — sur 4900). D'autre part, nous avons considéré comme positive l'émotion induite par une image chez un participant lorsque ce dernier lui a attribué une note supérieure à 5 (qui correspond sur l'échelle SAM à une valence neutre), et négative lorsque le spectateur lui a attribué une note inférieure à 5. Les émotions induites par une image auxquelles un participant avait attribué la valeur 5 n'étaient pas prises en compte dans l'analyse. Nous avons ensuite déterminé qu'il y

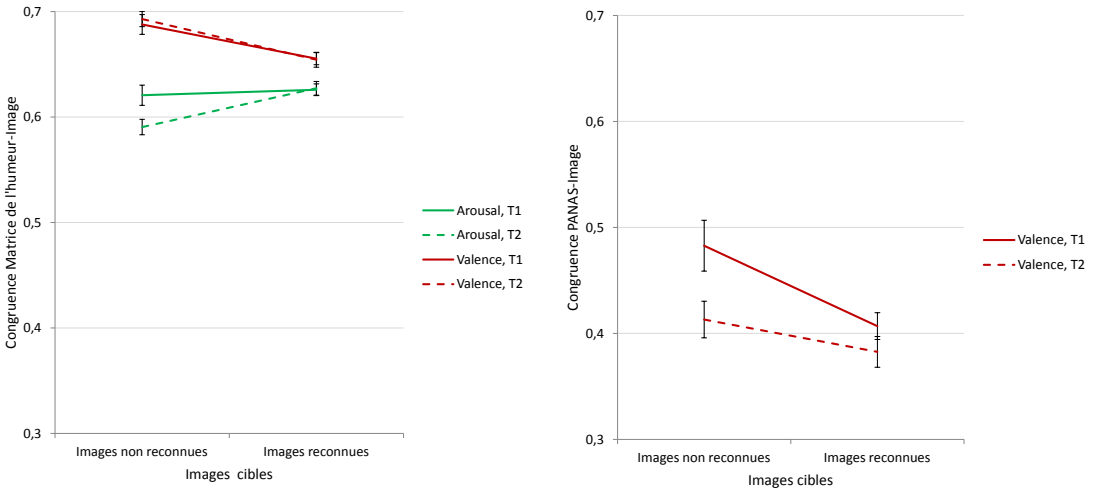


FIGURE 9.3 – Congruence émotionnelle moyenne entre les images répétées — reconnues *vs.* non reconnues — et l'état émotionnel du participant mesuré soit par la matrice de l'humeur (partie gauche), soit par l'I-PANAS-SF (partie droite). Plus les valeurs des ordonnées sont élevées, plus la congruence est forte (elle est minimum pour la valeur 0 et parfaite pour la valeur 1). Les résultats obtenus au premier test de mémoire (T1) sont matérialisés par des lignes continues, et ceux obtenus au second test de mémoire (T2) sont matérialisés par des lignes pointillées. Quant aux lignes vertes et rouges, elles représentent les résultats pour lesquels la congruence émotionnelle est mesurée sur les dimensions d'arousal et de valence, respectivement. Les barres d'erreur correspondent aux erreurs types des moyennes.

avait congruence émotionnelle lorsque l'état émotionnel du participant mesuré par l'I-PANAS-SF était d'une valence similaire — soit positive, soit négative — à l'émotion induite chez lui par l'image, telle qu'il l'avait évaluée (la variable prenait dans ce cas la valeur 1) ; dans le cas contraire, il n'y avait pas congruence (la variable prenait alors la valeur 0). Enfin, nous avons testé la moyenne de congruence émotionnelle (i.e. la moyenne des états 0 et 1 de la variable) en fonction du fait que l'image répétée avait été reconnue ou oubliée.

Les résultats sont présentés dans la partie droite de la figure 9.3. Un test U de Mann-Whitney indique que la congruence émotionnelle moyenne est significativement plus basse pour les images reconnues ($\mu = 0.407$) que pour les images non reconnues ($\mu = 0.483$) pour le premier test de mémoire ($Z = -2,822, p(= 0,00477) < .0083$). Un test similaire n'indique en revanche aucune différence significative similaire entre les images reconnues ($\mu = 0.383$) et les images non reconnues ($\mu = 0.413$) pour le second test de mémoire ($Z = -1,354, p = .1758$). Le résultat obtenu pour le premier test de mémoire va à l'encontre de notre hypothèse, puisque la congruence émotionnelle est plus faible pour les images non reconnues que reconnues. D'autre part, le fait que cet effet n'apparaisse pas pour le second test de mémoire doit encourager à la précaution. De nouvelles études sont nécessaires.

L'effet de facilitation de la récupération mnésique lorsque l'état émotionnel au moment de la récupération est similaire à l'état émotionnel au moment de l'encodage (Lewis and Critchley, 2003) a été montré pour la mémoire d'évènements autobiographiques, en utilisant l'hypothèse opérationnelle suivante : « La probabilité de récupérer un évènement est plus haute lorsque les états émotionnels lors de l'encodage et de la récupération concordent que lorsqu'ils ne concordent pas. » (Eich and Metcalfe, 1989, Eich et al., 1994). Nous avons cherché à mettre en lumière cet effet dans une tâche de reconnaissance d'images, avec l'hypothèse opérationnelle suivante : « La probabilité de reconnaître une image est d'autant plus haute que les états émotionnels lors de l'encodage et de la récupération coïncident fortement. » Il y a donc une différence de méthode considérable, et il y a des chances que l'effet n'existe pas dans une tâche de reconnaissance d'image. D'autant que cet effet n'a pas toujours été retrouvé, et que son apparition semble dépendre beaucoup de la méthodologie adoptée (voir à ce sujet la méta-analyse de (Ucros, 1989)). Cependant, il se peut aussi, comme on l'a dit, que nous n'ayons pas réussi à le mettre en valeur. Si cette seconde hypothèse est vraie, d'autres types d'analyses pourraient être envisagées. Il serait, par exemple, intéressant de réaliser des analyses de corrélation plus complexes (i.e. non linéaires). Plusieurs études sur la mémoire humaine suggèrent que nous nous souvenons des éléments d'un évènement en proportion de leur degré de distinctivité par rapport à leur contexte local (Eysenck, 2014, Nairne, 2006). En extrapolant, on pourrait émettre l'hypothèse que le contraste émotionnel produit par une image dans son contexte favorise sa reconnaissance. En somme, une image cible qui induit un état émotionnel soit proche de celui induit par

l'image qui la précède, soit éloigné, pourrait être mieux reconnue, ce qui supposerait une relation non linéaire entre la mémorabilité d'une image et l'interaction de l'émotion qu'elle véhicule avec son contexte émotionnel de présentation. Finalement, il serait intéressant de faire évaluer l'ensemble des images sur les dimensions d'arousal et de valence (i.e. pas uniquement les images cibles, mais également les images de remplissage) par l'ensemble des participants, au lieu de se baser sur un score d'émotion moyen, afin de gagner en précision. Même si l'information émotionnelle la plus susceptible d'être utilisée dans un futur proche pour la prédiction de la mémorabilité des images est celle qui peut être extraite algorithmiquement de l'image, qui renvoie à un score émotionnel moyen. Cependant, le développement d'outils pour mesurer en temps réel les émotions humaines progresse vite (nous en donnerons un exemple dans le chapitre 12), ouvrant la porte à la prise en compte de l'émotion ressentie par l'observateur particulier de l'image. Enfin, avoir un nombre plus important de données augmenterait également la probabilité de révéler un effet ténu, s'il existe.

9.5 Conclusion

L'étude présentée dans ce chapitre s'inscrit dans une volonté de prendre en compte les effets contextuels dans les modèles de prédiction de la mémorabilité des images ; en particulier, dans notre cas, pour ajuster les scores de mémorabilité calculés par MemoNet. Nous n'avons cependant pas obtenu de résultat concluant. Nous pensons toutefois qu'il demeure important de continuer à s'intéresser aux effets du contexte émotionnel sur la mémorabilité des images. En effet, il est probable que de tels effets existent. Le cas échéant, les avancées dans l'extraction computationnelle des émotions véhiculées par les images et dans les technologies de mesure en temps réel de l'état émotionnel permettront d'intégrer plus aisément de tels effets contextuels aux modèles de prédiction. Quant au contexte global, il existe de nombreux autres moyens de le prendre en compte, autant qu'on en peut imaginer ; par exemple, en calculant la distinctivité d'une image par rapport aux autres images sur la base de caractéristiques de bas niveau. C'est, comme nous l'avons précédemment évoqué, l'idée de Bylinskii *et al.*, qui signent une étude intéressante sur la modélisation des liens entre une telle distinctivité et la mémorabilité des images (Bylinskii et al., 2015b).

Le chapitre suivant traite des facteurs individuels, susceptibles d'influencer la mémorabilité des images.

10

Un modèle de l'influence des facteurs individuels sur la mémorabilité des images

Proposer la même image numérique à plusieurs individus n'est pas forcément leur proposer la même expérience. Dans ce chapitre, nous nous intéressons à la prise en compte des facteurs individuels qui influencent l'expérience suscitée par une image et sa mémorabilité. Dans un premier temps, nous nous focalisons sur les influences comparées de la personnalité — plutôt masculine ou féminine — et du sexe biologique d'un individu sur son expérience de la négativité d'une image. Ensuite, nous proposons un modèle de régression logistique qui inclut plusieurs facteurs individuels qui influencent significativement la probabilité qu'une image soit mémorisée. Nous comparons également ce modèle *white box* à un modèle *black box* basé sur un séparateur à vaste marge. Notre objectif serait d'utiliser un modèle semblable à celui proposé dans ce chapitre pour adapter à chaque individu les scores de mémorabilité moyens générés pour les images par les modèles de prédiction de la mémorabilité, tels que MemoNet.

10.1 Introduction

Nous inscrivons le travail sur la mémorabilité des images présenté dans ce chapitre dans le cadre de la qualité d'expérience. Une communauté forte de nombreux chercheurs travaille dans ce champ de recherche, pour lesquels fournir grâce à un contenu numé-

rique une expérience excellente représente une sorte de Graal. La personnalisation des contenus numériques proposés aux utilisateurs, souvent laissée de côté à cause de la difficulté à prendre en compte les facteurs individuels, pourrait s'avérer utile, ou même nécessaire, pour atteindre cette excellence. La manière de prendre en compte l'individu dans une mesure de l'importance d'une expérience suscitée par un contenu numérique — c'est-à-dire de sa mémorabilité — pourrait, par conséquent, intéresser les chercheurs en qualité d'expérience.

10.1.1 Une étude intéressante en qualité d'expérience

Lorsqu'il est question de distribuer un contenu multimédia à un utilisateur — dans notre cas, une image numérique —, l'aspect technique nous vient d'abord en tête : comment produire, transmettre, afficher le contenu de manière qu'il soit, en bout de chaîne, consommé par l'utilisateur dans sa meilleure forme. Cependant, même si toutes ces étapes sont réalisées dans les règles de l'art, cela ne garantit pas à l'utilisateur la meilleure expérience possible. En effet, au-delà des considérations techniques, il y a un utilisateur, qui ressent, pense, juge — qui fait l'expérience du contenu.

Proposer une expérience — excellente — à un utilisateur, plutôt qu'un simple contenu multimédia, représente un objectif suprême pour de nombreux chercheurs travaillant dans les technologies du multimédia. En sorte qu'un nouveau champ de recherche est apparu durant la décennie précédente : l'imagerie numérique s'est tournée vers l'expérience utilisateur. Dans ce champ d'investigation, l'accent est mis sur l'éventail des réactions humaines — attentionnelles, émotionnelles, mnésiques, etc. — impliquées dans l'évaluation de la qualité d'image. Le point de convergence principal des attentions s'est déplacé de l'image vers l'utilisateur.

En raison de la difficulté à mesurer de façon non intrusive et à intégrer aux modèles computationnels de qualité des images les différences interindividuelles, les expériences multimédias sont optimisées pour un utilisateur *moyen*. La qualité de telles expériences est évaluée à l'aide de scores d'opinion moyens (Union,). Il y a donc une perte des différences interindividuelles, c'est-à-dire une perte de l'information nécessaire à la personnalisation des expériences proposées. La prise en compte de l'observateur nécessiterait une mise en lumière et une modélisation des facteurs individuels qui influencent l'expérience suscitée par un contenu multimédia.

A travers la description d'une étude portant sur les facteurs individuels influençant la mémorabilité de l'expérience suscitée par une image numérique, ce chapitre, en plus d'être intéressant dans le cadre de la prédiction de la mémorabilité d'images, le sera plus largement pour les chercheurs en qualité d'expérience.

10.1.2 De la qualité de l'image à la pertinence de l'image

Le nombre — toujours plus — impressionnant d'images numériques partagées chaque jour grâce à internet, rend extrêmement souhaitable la possibilité de prédire automatiquement leur pertinence pour l'utilisateur à qui elles sont destinées. Des centaines de millions d'images inédites sont téléversées chaque jour sur Facebook. Google assurait, en 2008, avoir indexé plus de 1000 milliards d'images.

Prenons une base d'images constituée de millions d'exemplaires. Nous nous somme posés la question suivante : comment choisir l'image qui est la plus pertinente pour un utilisateur particulier — dans le sens où elle suscitera chez lui la meilleure expérience possible ? Nous pourrions répondre : choisissons l'image qui a la meilleure qualité. Prenons l'exemple suivant. En 2010, puis en 2011, Facebook a permis de stocker des images de plus haute résolution. La progression est illustré par la Figure 10.1. Si l'on s'intéresse au partage des images pendant cette période, à l'instar de (Fedorovskaya and De Ridder, 2013), on peut se rendre compte qu'alors que plus de trois milliards d'images ont été téléchargées chaque mois les utilisateurs ont partagé et affiché les images indépendamment de leur résolution et de leur qualité. Cela signifie que les utilisateurs sont prêts à laisser de côté la qualité pour quelque chose qui a, à leurs yeux, plus de valeur. Cet exemple montre que proposer aux fournisseurs et aux utilisateurs d'images numériques des aides pour la sélection et le partage des images qui s'appuient uniquement sur l'utilisation de métriques objectives de qualité d'images (i.e. des algorithmes qui déterminent automatiquement la qualité des images) n'est pas toujours pertinent. La question qui en découle est la suivante : comment développer une mesure objective de la pertinence d'une image pour un utilisateur particulier, en s'appuyant sur l'importance de l'expérience qu'elle suscitera chez celui-ci ?

Il a récemment été proposé que la mémorabilité était un indicateur pertinent pour choisir les images les plus pertinentes d'une collection d'images ou de vidéos (p. ex. pour extraire l'image-clé qui présentera la vidéo) (Isola et al., 2011b). Comme l'esthétique d'une image, son intérêt, et d'autres métriques qui peuvent être liées à la pertinence d'une image, la mémorabilité nous renseigne de l'utilité d'une image dans la vie quotidienne. La mémorabilité d'une image peut être, dans une certaine mesure, prédite computationnellement (Isola et al., 2014). Cependant, comme nous l'avons expliqué dans le chapitre 4, les modèles de prédiction existants se basent presqu'exclusivement sur les informations intrinsèques des images : à notre connaissance, aucun ne prend en compte l'observateur particulier de l'image. Or, proposer la même image à plusieurs utilisateurs ne revient pas nécessairement à leur proposer la même expérience. En conséquence, la mémorabilité d'une image, puisqu'elle dépend de l'expérience de l'utilisateur, change dans une certaine mesure avec l'individu et le contexte. L'intégration de l'idiosyncrasie dans les modèles de prédiction de la mémorabilité pourrait donc améliorer leur performance. Pour atteindre cet objectif, des études sont nécessaires pour mieux comprendre les facteurs individuels qui influencent l'expérience induite par une image et, par suite,

FIGURE 10.1 – Augmentation progressive de la taille maximum des images sur Facebook. Les utilisateurs partagent et affichent les images indépendamment de leur résolution et de leur qualité. (Tiré de (Fedorovskaya and De Ridder, 2013).)

sa mémorabilité, et pour modéliser ces facteurs. La mémorabilité étant un indicateur de l'importance d'une image, ces facteurs seraient intéressants pour les systèmes d'optimisation de l'expérience utilisateur, qui pourraient alors personnaliser les contenus numériques proposés aux utilisateurs.

D'autre part, comme nous l'avons précédemment établi, l'émotion qu'une image véhicule est étroitement liée à sa mémorabilité. Cependant, les études portant sur la mémorabilité des images en vision par ordinateur n'ont, à notre connaissance, ne se sont pas intéressées aux émotions véhiculées par les images. nous avons expliqué qu'il était possible d'extraire automatiquement d'une image de l'information émotionnelle (Liu et al., 2010b, Lucassen et al., 2010, Wei et al., 2008, Gbèhounou et al., 2012, Machajdik and Hanbury, 2010, Ou et al., 2004), et que cette information pourrait être utilisée pour améliorer la précision des scores de mémorabilité générés par les modèles computationnels de prédiction. Cependant, un tel objectif requerrait une modélisation préalable des liens entre l'émotion véhiculée par une image et sa mémorabilité. Dans le chapitre 7, nous avons étudié de tels liens à travers l'étude des scores moyens d'émotion et de mémorabilité des images de notre base de données. Il serait intéressant, puisque l'émotion suscitée par une image, comme sa mémorabilité, sont en partir subjectives, d'étudier ces mêmes liens en prenant en compte les individus. Une telle étude peut être réalisée à partir des données individuelles de mémorabilité, d'arousal et de valence que nous avons collectées.

10.1.3 Les facteurs individuels qui pourraient influencer la mémorabilité

Comme nous allons le voir dans cette section, plusieurs facteurs individuels ont été isolés par les chercheurs en psychologie pour leur influence à la fois sur l'expérience émotionnelle induite chez l'observateur par une image et sur la mémorabilité de l'image, incluant l'âge, le genre, l'état affectif (à la fois durable ou *trait* et actuel ou *état*) et la fatigue.

Quelques études ont montré que l'âge influençait la notation de l'arousal et de la valence suscités par une image. Par exemple, Grühn et Scheibe ont montré que les adultes âgés évaluaient les images négatives comme plus négatives et les images positives comme plus positives que les jeunes adultes, et qu'ils évaluaient également comme plus élevé l'arousal suscité par les images négatives, et plus bas l'arousal suscité par les images positives, que les jeunes adultes (Grühn and Scheibe, 2008). De plus, le fait, bien documenté, que la mémoire est généralement meilleure pour les images négatives semble être moins évident chez les adultes âgés que chez les jeunes adultes (possiblement à cause de l'amélioration de la régulation des émotions avec l'âge) (Charles et al., 2003, Mikels et al., 2005).

La littérature suggère également des différences dans l'évaluation de l'émotion suscitée par les images entre les hommes et les femmes. En particulier, il semble que les femmes ressentent plus intensément la négativité des images que les hommes. Par exemple, Wrase *et al.* ont observé, dans une étude en neuroimagerie, que les femmes montraient une activation cérébrale plus forte dans les gyri cingulaires antérieur et médian (qui sont parties intégrantes du système limbique, impliqué dans la formation des émotions — p. ex. (Hadland et al., 2003) — et la mémoire — p. ex. (Kozlovskiy et al., 2012)) que les hommes lorsqu'elles regardaient des images négatives (Wrase et al., 2003). Dans une autre étude, de neurophysiologie, Lithari *et al.* ont également montré que les amplitudes des ondes correspondant aux potentiels évoqués (i.e. à la modification du potentiel électrique produite par le système nerveux en réponse aux stimulations) tendaient à être plus fortes chez les femmes que chez les hommes lors du visionnage de stimuli négatifs (Lithari et al., 2009).

Ces derniers résultats paraissent suggérer que le sexe biologique détermine le fait d'être plus ou moins sensible aux stimuli négatifs. Cependant, cette différence de sensibilité pourrait n'être pas directement liée au sexe biologique, mais à la personnalité des individus : il suffit pour cela que la probabilité de développer une personnalité plus sensible à la négativité que la moyenne soit plus forte chez les femmes que chez les hommes, et qu'elle soit influencée (au moins en partie) par des facteurs développementaux. Par exemple, il est probable que le développement de la personnalité soit lié aux rôles de genre attendus par la société (Eccles et al., 1990). La notion de rôle de genre renvoie à la tendance des individus à intégrer des comportements appropriés aux attentes

sociales liées à leur sexe. Plusieurs questionnaires ont été développés pour étudier les rôles de genre, le plus connu d'entre eux étant le BSRI (pour *Bem Sex Role Inventory*) (Bem, 1977, Bem, 1981). Ce questionnaire, que nous avons utilisé dans l'expérience décrite au chapitre 5, évalue la façon dont les individus s'identifient aux rôles de genre, et mesure en particulier la masculinité-féminité. En utilisant les résultats à ce questionnaire, il nous sera possible de déterminer si la personnalité dominante (masculine ou féminine) ou le sexe biologique (homme ou femme) a la plus grande influence sur l'évaluation de l'émotion suscitée par une image.

L'état émotionnel d'un individu peut interagir avec l'émotion suscitée par un stimulus. Un exemple est l'effet de congruence émotionnel (voir la section 3.1.3 du chapitre 3). L'existence de ce biais a notamment été montré dans la dépression : une humeur dépressive augmente la probabilité de se souvenir d'évènements négatifs (p. ex. (Watkins et al., 1992, Drace, 2013)). On parlera d'émotion *trait* pour désigner cette humeur, dont la durée est de loin supérieure à celle de la tâche. En extrapolant, l'émotion *état*, qui renvoie à l'humeur passagère d'un individu, pourrait, au moment celui-ci réalise une tâche de mémoire, interagir avec l'état psychologique modifié par un stimulus.

Enfin, il est connu depuis longtemps que la fatigue tend à diminuer les performances cognitive en général (dont la mémoire) et affecte l'état émotionnel. Pour prendre un exemple extrême, la privation de sommeil est connue pour avoir un effet critique à la fois sur l'humeur et la mémoire (Hart et al., 1987).

10.1.4 Objectif de l'étude

Pour mieux comprendre les facteurs idiosyncratiques sous-jacents qui influencent l'expérience émotionnelle suscitée par une image et sa mémorabilité, dans l'expérience décrite au chapitre 5, nous avons collecté des informations sur les participants. Préalablement à la session de test, les participants répondaient à des questions et questionnaires d'auto-évaluation, portant sur : leurs âge, genre, état émotionnel, état de fatigue et traits de personnalité (BSRI). Ensuite, les ils passaient les deux tests de mémoire, puis évaluaient les images sur les dimensions d'arousal et de valence. Sur la base de ces données, dans la suite de ce chapitre nous effectuons une analyse comparative des influences du sexe biologiques *versus* de la personnalité dominante mesurée par le BSRI sur l'évaluation de la valence induite par des images. Ensuite, nous dérivons des données un modèle des liens entre la mémorabilité des images et les facteurs individuels mesurés.

10.2 Rappel de la méthode employée et données utilisées

La méthode pour collecter les données sur les participants a été décrite dans le chapitre 5. Les deux paragraphes suivants sont un simple rappel, qui ne concerne que les

informations utilisées dans ce chapitre.

Nous avons utilisé le BSRI (Bem, 1981) pour mesurer la masculinité-féminité des participants. Ce questionnaire auto-administré est composé de 60 items, pour chacun desquels la personne qui s'évalue doit noter à l'aide d'une échelle en sept points (1 : « jamais ou presque jamais vrai » ; 7 : « toujours ou presque toujours vrai ») dans quelle mesure il est adapté pour décrire sa personnalité. Sur la base des réponses du participant, le BSRI permet de catégoriser sa personnalité comme étant globalement « masculine », « féminine », ou « indifférenciée » (i.e. que les pôles masculin et féminin sont sous-investis). L'I-PANAS-SF (Thompson, 2007) était utilisé pour mesuré l'état affectif dominant du participant au cours de la dernière année. L'I-PANAS-SF fournit des résultats sur deux dimensions : la positivité de l'état affectif, et sa négativité. La matrice de l'humeur proposée par Eich et Metcalfe (Eich and Metcalfe, 1989) était utilisée pour évaluer l'humeur des participants au moment de la tâche réalisée. Elle fournit une valeur de valence et une valeur d'arousal, chacune entre 1 (i.e. le degré le plus bas) et 9 (i.e. le degré le plus haut). Nous avons utilisé un unique item pour mesurer le niveau de fatigue des participants (i.e. « Indiquez dans quelle mesure vous vous sentez fatigué. « 1 » très fatigué « ; 5 : « absolument reposé »). Enfin, nous demandions aux participants de renseigner leur âge et leur genre.

Les émotions suscitées par les images étaient évaluées grâce aux échelles SAM à neuf degrés (Bradley and Lang, 1994). La performance de mémoire était mesurée quelques minutes après l'encodage mnésique, puis un jour après. Le modèle présenté dans la suite de ce chapitre porte sur les scores de mémorabilité calculés à partir des mesures de mémoire effectuées un jour après l'encodage, plus proches, comme nous l'avons précédemment expliqué, d'une mémorabilité *durable*, qui répond mieux aux attentes des chercheurs travaillant sur la mémorabilité des images en vision par ordinateur. Les données collectées à l'aide de la matrice de l'humeur et de la question de fatigue que nous avons utilisées sont celles qui correspondent aux réponses apportées à ces deux outils par les participants juste avant ce second test de mémoire.

10.3 Résultats

Les résultats sont présentés en deux parties. Premièrement, nous testons l'hypothèse selon laquelle la personnalité dominante (masculine ou féminine) a une influence plus importante que le sexe biologique sur la manière dont les individus évaluent la valence des émotions induites par des images. Ensuite, nous dérivons un modèle mathématique des données collectées, qui quantifie l'influence des facteurs individuels sur la probabilité qu'une image soit reconnue lorsqu'elle est répétée plutôt qu'oubliée.

10.3.1 Catégorisation sur la base des scores au BSRI

L'analyse des résultats du BSRI nécessite un pré-traitement des données brutes. Pour ce pré-traitement, nous avons utilisé une technique basée sur l'utilisation d'une médiane calculée à partir de nos données, recommandée par Bem en 1977 (Bem, 1977) à la suite des critiques de Spence *et al.* portant sur sa procédure de traitement originelle, qui ne permet pas de distinguer les individus ayant des scores élevées, ou ceux ayant des scores faibles, à la fois sur les items « masculins » et « féminins » (Spence et al., 1975). Le BSRI sépare les individus en quatre catégories de personnalité dominante : « masculine », « féminine », « androgyne » et « indifférenciée ». Une classification typique (« masculine » ou « féminine ») est le résultat d'une notation au-dessus de la médiane pour les items associés à un des deux genres, et en-dessous de la médiane pour les items associés à l'autre genre. Lorsqu'un individu a noté au-dessus de la médiane à fois les items masculins et féminins, il est rangé dans la catégorie « androgyne ». Lorsqu'un individu a noté au-dessous de la médiane à fois les items masculins et féminins, il est rangé dans la catégorie « indifférenciée » (qui signifierait donc qu'il possède très peu des traits masculins ou féminins tels que la société les définit).

Les résultats de la catégorisation en fonction du sexe biologique des participants sont présentés dans le tableau 10.1. Nous pouvons y voir que 48% des hommes ont été classés dans la catégorie « masculine » et 24% dans la catégorie « féminine », tandis que 57% des femmes ont été classées dans la catégorie « féminine », et 32% dans la catégorie « masculine ».

TABLE 10.1 – Ventilation des participants en fonction de leur sexe biologique dans les catégories proposées par le BSRI.

	Hommes	Femmes
Masculine	10	9
Féminine	5	16
Androgyne	4	2
Indifférenciée	2	1

10.3.2 Effets du sexe biologique et de la personnalité dominante sur la notation de la valence

Nous avons effectué une ANOVA à deux facteurs pour plan non équilibré pour tester les effets du sexe biologique et de la personnalité dominante telle que déterminée par le BSRI (avec seulement les deux catégories « masculine » et « féminine »)

sur l'évaluation de la valence des images (voir la figure 10.2). Nous n'avons pas observé d'effet significatif du sexe biologique (homme, femme) sur la notation de la valence ($F(1) = 0.22, p = .641$). En revanche, nous avons observé un effet significatif de la personnalité dominante (masculine, féminine) sur la notation de la valence ($F(1) = 5.29, p < .001$) : les individus classés dans la catégorie « féminine » évaluaient en moyenne les images comme plus négatives que les individus classés dans la catégorie « masculine ».

Comme la littérature laisse à penser que les femmes évaluent les images négatives comme plus négatives que les hommes ne les évaluent, en raison de leur plus grande réceptivité aux stimuli négatifs (Wrase et al., 2003, Lithari et al., 2009), nous avons reproduit la même analyse en utilisant seulement les images négatives (i.e. les images pour lesquelles le score de valence moyen était inférieur à la valeur 5). Nous avons observé un pattern similaire, avec aucun effet significatif du sexe biologique sur la notation de la valence de l'expérience suscitée par les images négatives ($F(1) = 0.67, p = .412$), mais un effet significatif de la personnalité dominante (masculine ou féminine) des participants sur cette même notation ($F(1) = 15.85, p < .001$).

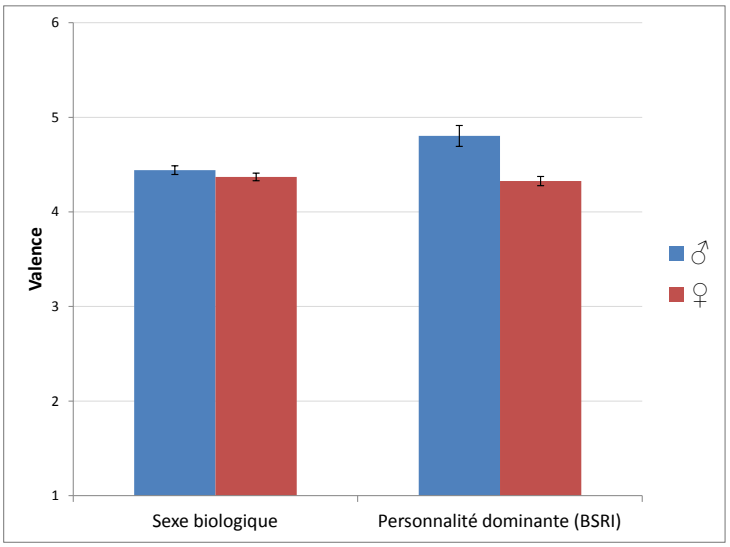


FIGURE 10.2 – Notation moyenne de la valence de l'expérience émotionnelle provoquée par une image en fonction du sexe biologique des participants et de leur personnalité dominante (masculine ou féminine) telle que déterminée par le BSRI. (Les barres d'erreur correspondent aux erreurs-types associées aux moyennes.)

10.3.3 Modèle *white box*

Pour modéliser l'effet des facteurs individuels sur la détection de la répétition d'une image cible dans le second test de mémoire (i.e. dans la tâche de reconnaissance pas-

sée un jour après l'encodage mnésique), nous avons effectué une régression logistique, en utilisant les facteurs individuels comme régresseurs et la détection de la répétition d'une image cible (avec deux résultats discrets possibles, « détectée » ou « non détectée ») comme variable binomiale. Nous avons utilisé les prédicteurs suivants : la notation de la valence de l'image par le participant, son âge, son sexe biologique, sa personnalité dominante telle que déterminée par le BSRI (avec les quatre modalités possibles, « masculine », « féminine », « androgyne » et « indifférenciée »), l'état émotionnel durable mesuré par l'I-PANAS-SF (sur les dimensions « positive » et « négative »), l'état émotionnel au moment de la tâche mesuré par la matrice de l'humeur (sur les dimensions d'arousal et de valence) et le niveau de fatigue.

Le modèle, dans le cas d'une seule observation, peut être écrit comme suit :

$$y_n = \beta_0 + \sum_{k=1}^{K} x_{nk}\beta_k + \epsilon_n \tag{10.1}$$

où y est la variable dépendante — le succès ou l'échec de la détection de la répétition de l'image —, x les valeurs de nos prédicteurs, β les coefficients à estimer, et ϵ le terme d'erreur. Même si nous supposions que les relations entre les facteurs individuels susmentionnés et notre variable binomiale étaient plus complexes que des relations linéaires, nous nous sommes cantonnés à la modélisation de ce dernier type de relation afin d'obtenir un résultat interprétable.

Les résultats de la régression logistique sont présentés dans la figure 10.3. Chacun des coefficients calculés exprime l'influence d'un facteur individuel sur la chance relative que la répétition d'une image soit détectée plutôt que manquée. La méthode employée donne également une mesure de la significativité statistique des prédicteurs du modèle en leurs associant une $p - value$.

Selon le modèle obtenu, plus l'expérience émotionnelle suscitée par l'image est positive, moins elle est susceptible d'être reconnue. Au contraire, plus l'état émotionnel *trait* (i.e. durable) est positif, plus une image est susceptible d'être reconnue ; et plus cet état émotionnel est négatif, moins une image a de chance d'être reconnue. Enfin, plus le niveau d'arousal rapporté par le participant à propos de son humeur au moment de la tâche est élevé, plus une image a de chance d'être reconnue. Nous n'avons trouvé aucun effet significatif de l'âge, du sexe biologique, de la personnalité dominante, de la valence de l'humeur au moment de la tâche ou du niveau de fatigue sur la probabilité qu'une image soit reconnue.

Pour tester notre modèle mathématique, nous avons employé une technique de validation croisée reposant sur des sous-échantillonnages aléatoires (avec 75% des données — soit 3875 observations — utilisées pour l'apprentissage, et 25% pour la classification, pour chacun des 40 fractionnements effectués). Notre modèle obtient un taux de classification correcte de 66%, avec un intervalle de confiance de 3%.

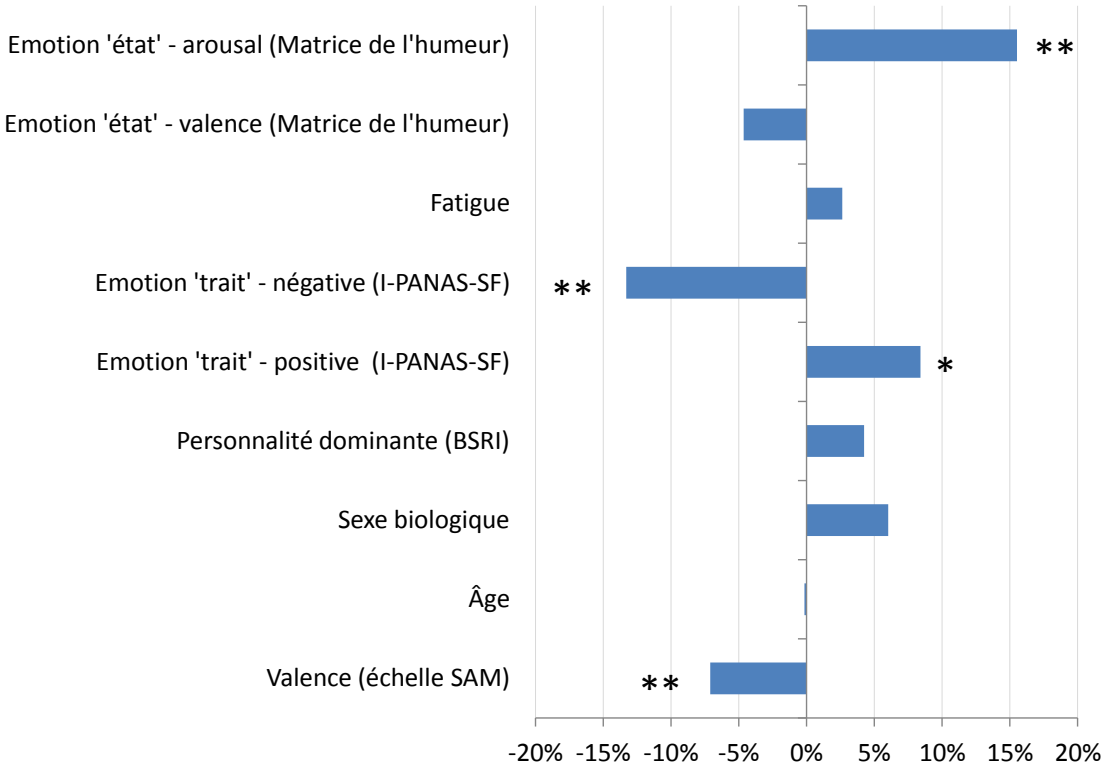


FIGURE 10.3 – Coefficients de régression calculés pour chacun des facteurs individuels, qui expriment la relation linéaire de chaque facteur avec la probabilité qu'une image répétée soit détectée plutôt que manquée. ($*p < .05$, $**p < .001$).

10.3.4 Modèle *black box*

La méthode de régression logistique ne permet de modéliser que des relations simples (i.e. linéaires) entre les variables. L'avantage est que le modèle obtenu, *white box*, est aisément interprétable ; l'inconvénient est qu'il *manque* les interactions complexes qui peuvent exister entre les variables. Comme cela a été souligné par (Mohammadi and Vinciarelli, 2012), il peut résulter du choix d'utiliser une méthode de régression logistique pour la modélisation un taux de classification du modèle obtenu assez faible, de sorte qu'il peut s'avérer intéressant de compléter l'analyse de données en utilisant un séparateur à vaste marge (SVM), qui est capable de modéliser des relations plus complexes entre les facteurs. L'inconvénient du modèle, *black box*, obtenu par une telle méthode, est que les fonctions mathématiques reliant les entrées aux sorties du modèle sont difficilement interprétables.

Dans cette étude, nous avons entraîné un SVM pour obtenir un point de comparaison pour juger de l'efficacité de notre modèle. Nous avons opté pour un SVM à noyau linéaire (qui permet à un classifieur linéaire de résoudre un problème non linéaire en ajoutant des dimensions à l'espace de représentation des données d'entrée, dimensions dans lesquelles les données sont linéairement séparables) comme classifieur, et utilisé les mêmes données et la même méthode de validation croisée que pour le modèle de régression logistique. Nous avons obtenus un taux de classification correcte de 71%, avec un intervalle de confiance de 3%. Cette performance nous donne une idée du degré d'optimalité de notre modèle de régression logistique : en perdant un peu de performance en utilisant un tel modèle plutôt qu'un SVM, nous gagnons néanmoins la possibilité d'interpréter les relations entrées-sorties.

10.4 Discussion

Effets comparés du sexe biologique et de la personnalité dominante sur l'évaluation de la valence

Nous n'avons pas observé de différence significative entre les hommes et les femmes relativement à leur évaluation de la valence des images. Au contraire, nous avons observé une différence significative entre les individus possédant une personnalité à dominante masculine et ceux possédant une personnalité à dominante féminine, avec les individus de la catégorie « féminine » évaluant en moyenne les expériences émotionnelles suscitées par les images comme plus négatives que les individus de la catégorie « masculine ». Ce résultat suggère que la personnalité a une influence plus importante que le sexe biologique sur la manière dont les individus font l'expérience de la positivité/négativité des émotions. À notre connaissance, c'est la première qu'un tel résultat est rapporté ; de nouvelles études sont donc nécessaires pour le confirmer.

Nous avons observé le même pattern (i.e. pas d'effet du sexe biologique) pour les images négatives uniquement, alors que la littérature suggère que les femmes ressentent en général avec plus d'intensité les expériences négatives (Lithari et al., 2009, Wrase et al., 2003). Cette absence de résultat significatif semble procéder d'un problème de puissance, puisque nous avons trouvé que les hommes ont plus souvent une personnalité masculine, et les femmes une personnalité féminine, et que la masculinité-féminité a un impact sur la notation de la valence. La manière dont nous avons mesuré la valence de l'expérience induite par les images — c'est-à-dire à l'aide d'outil d'auto-évaluation — pourrait expliquer ce manque de puissance. Les études mentionnées avaient recours à des techniques de mesure moins intrusives, à savoir l'électroencéphalographie pour mesurer les potentiels évoqués ainsi qu'un capteur de conductance cutanée pour (Lithari et al., 2009), et l'IRM fonctionnelle pour (Wrase et al., 2003).

Plus généralement, la catégorisation de la personnalité des individus à l'aide du BSRI doit être considérée avec prudence : entre autres, les items correspondent à des stéréotypes de genre dans la société américaine des années 70, et leur généralisation à la société française actuelle peut être discutée.

Modélisation

Dans notre modèle de régression, plus l'expérience induite par l'image est évaluée positivement par un individu, plus la probabilité qu'il reconnaisse cette image lorsqu'elle est répétée est faible. On peut rapprocher ce résultat de celui, similaire, présenté dans le chapitre 7, qui porte sur l'étude des scores de mémorabilité et d'émotion moyens des images : les images négatives étaient mieux reconnues que les images neutres et positives. Comme nous l'avons évoqué dans ce chapitre, ainsi que dans la section 3.2.1 du chapitre 3, de nombreuses études ont montré que les informations positives et négatives sont mieux mémorisées que les informations neutres ; cela suggère une relation non linéaire entre la valence et la mémorabilité des images. Le résultat obtenu concorde avec ceux de plusieurs études portant sur de jeunes adultes, qui ont trouvé que l'information négative était mieux mémorisée que l'information positive (Dewhurst and Parry, 2000, Mather et al., 2000, Ochsner, 2000). Ce biais de négativité (qui, dans cette acception, réfère à la propension à mémoriser plus facilement les informations négatives que positives) pourrait s'atténuer, voir disparaître avec l'âge, comme cela est suggéré par Charles *et al.*, qui ont observé que le nombre d'images négatives relativement au nombre d'images positives et neutres rappelées baissait avec l'âge (Charles et al., 2003). A ce propos, il est important de noter que nos participants étaient de jeunes adultes (le plus âgé d'entre eux avait 41 ans).

Lorsque la négativité de l'état émotionnel *trait* d'un individu augmente dans notre modèle, la probabilité qu'une image soit reconnue diminue ; au contraire, lorsque la positivité de l'état émotionnel *trait* augmente, cette probabilité augmente. Cela suggère qu'un état émotionnel durablement négatif a un impact négatif sur la mémoire, alors

qu'un état émotionnel durablement positif a un impact positif sur la mémoire. Ce résultat s'accorde avec le fait que la dépression est souvent associée à une mémoire globalement défaillante (pour une revue sur la question, voir (Burt et al., 1995)).

Selon notre modèle, plus l'arousal de l'humeur d'un individu est élevé au moment où il passe la tâche, plus la probabilité qu'une image soit reconnue augmente ; la valence de son humeur n'influence cependant pas significativement cette probabilité. Ces résultats suggèrent que plus un individu est éveillé ou stimulé, plus sa performance de mémoire est bonne, alors que la positivité/négativité de son état émotionnel transitoire n'influence pas, ou moins, cette performance. Il est intéressant d'évoquer ici la notion de variable confondue : l'attention et le traitement de l'information, dont on a précédemment dit qu'ils étaient influencés par l'arousal et qu'ils influençaient à leur tour la mémoire (Christianson, 2014), ont pu agir comme des variables intermédiaires entre ces deux phénomènes psychologiques.

Peut-être parce que les relations de ces facteurs avec la performance de reconnaissance des images ne sont pas linéaires, ou à cause d'un nombre trop faible de participants dans notre étude, les prédicteurs de notre modèle que sont l'âge, le sexe biologique, la personnalité mesurée par le BSRI et le niveau de fatigue, n'influencent pas significativement la probabilité qu'une image soit reconnue. Concernant l'âge, notre groupe de participants consistait seulement en des adultes jeunes. Or, la performance de mémoire décline avec le vieillissement normal (par opposition au vieillissement pathologique) (Hedden and Gabrieli, 2004), mais ce phénomène advenant généralement assez tardivement, nous n'aurons pas pu le mettre en lumière. De plus, le biais de négativité tend, ainsi que nous l'avons exposé, à s'estomper avec l'âge, et nous n'aurons également pas pu en mesurer l'effet. Nous sommes donc potentiellement passé à côté de deux effets lié au facteur Âge. Il serait intéressant, pour une base d'images destinée à l'étude de la mémorabilité des images, que les scores de mémorabilité des images qui la constituent prennent en compte des mesures de performance de mémoire d'adultes âgés. Concernant le sexe biologique et la personnalité dominante, la littérature suggère que le sexe biologique influence la valence induite par une image, et nos résultats suggèrent que la personnalité dominante influence cette valence. Or, la valence et la mémorabilité des images sont liées. Peut-être existe-t-il donc un effet indirect du sexe biologique et de la personnalité dominante sur la mémorabilité des images, qui s'exerce par l'intermédiaire de la valence. Le fait qu'un tel effet serait indirect suggère qu'il serait également plus ténu et plus complexe qu'un effet direct ; notre méthode d'analyse aura, dès lors, était trop peu puissante pour le révéler. Quant à l'effet de la fatigue ressentie, il semble d'après nos résultats qu'une fatigue modérée (seuls trois participants ont rapporté être fatigués ou très fatigués — i.e. 1 ou 2 sur l'échelle associée à l'item de fatigue) n'a pas d'effet critique sur la mémoire, tel qu'une grande fatigue pourrait en avoir (Hart et al., 1987).

10.5 Conclusion

Dans ce chapitre, nous avons proposé un modèle de régression logistique pour expliquer la probabilité de reconnaître une image en fonction de différents facteurs individuels. Nous proposons d'introduire ce type de modèle dans le cadre de la prédiction de la mémorabilité des images, comme un outil pour personnaliser la mémorabilité prédite pour les images par les modèles existants. Dans un tel objectif, il suffira d'ajuster les scores de mémorabilité moyens générés par les modèles prédictifs à l'observateur particulier de l'image, sur la base des informations dont on dispose sur lui, rendues exploitables par notre modèle.

La modélisation des liens entre les facteurs utilisateurs et la mémorabilité est d'autant plus intéressante que l'acquisition de données individuelles est en plein essor, mue par l'intérêt croissant porté aux recherches liées aux interactions homme-machine (portant sur le *Quantified Self*, les interfaces neuronales directes, l'informatique affective, etc.). En particulier, — puisque les facteurs que nous avons mis en lumière sont liés à l'état émotionnel des individus — l'acquisition informatique de données émotionnelles a connu ces dernières années une impulsion considérable, avec la montée en puissance de l'informatique affective (qui a une revue IEEE spécialement dédiée depuis 2010 : *IEEE Transactions on Affective Computing*) (Picard, 2010). Les avancées rapides dans ces champs de recherche promettent pour un futur proche une diffusion à un large public d'outils d'acquisition de données individuelles (p. ex. casques d'EEG, oculomètres, capteurs de battements cardiaques, etc.). Certains de ces outils sont d'ores et déjà disponible ; par exemple, le dispositif EEG d'Emotiv (Emotiv,) que nous utilisons dans l'expérience décrite au chapitre 12, qui permet de mesurer et d'interpréter en temps réel l'état émotionnel d'individus. Pour mettre à profit de telles avancées technologiques dans le cadre de la prédiction de la mémorabilité, il est nécessaire que des études se concentrent sur la manière dont peuvent être intégrés des données individuelles aux systèmes de prédiction. La contribution de ce chapitre se veut une première pierre à cet édifice.

La manière de prendre en compte l'idiosyncrasie pour déterminer la mémorabilité d'une expérience émotionnelle (induite par un contenu numérique) présentée dans ce chapitre pourrait s'avérer particulièrement intéressante pour les chercheurs dont les travaux portent sur la qualité d'expérience. La mémorabilité peut, en effet, être considérée comme un indicateur de la pertinence de l'expérience proposée à un utilisateur par l'intermédiaire d'un contenu numérique. La prise en compte des facteurs individuels pour maximiser la mémorabilité des contenus numériques proposés à des utilisateurs est donc susceptible d'intéresser ces chercheurs. Au-delà de la mémorabilité, l'intégration des facteurs utilisateurs dans les systèmes d'optimisation de l'expérience, dans un objectif de personnalisation de la distribution des contenus numériques, est susceptible d'intéresser particulièrement les distributeurs de contenus.

Conclusion

Nous avons étrenné l'apprentissage profond pour la prédiction de la mémorabilité des images : notre modèle, MemoNet, obtient les meilleurs résultats à ce jour. Nous avons ensuite cherché de nouvelles voies pour améliorer la précision des prédictions, en considérant les informations extrinsèques — contextuelles et individuelles — des images potentiellement disponibles pour les modèles.

Probablement pour une ou plusieurs des raisons dont nous avons discuté dans la section consacrée, nos analyses n'ont pas révélé d'effets contextuels sur la mémorabilité des images. Loin de nous, cependant, l'idée d'abandonner cette voie. Bylinskii *et al.* ont d'ailleurs récemment obtenu des résultats engageant en prenant en compte le contexte de présentation des images dans la prédiction de leur mémorabilité (Bylinskii et al., 2015b). De nouvelles études seraient bienvenues, pour mesurer et révéler des effets contextuels, les modéliser, et les intégrer aux modèles de prédiction.

Quant à la prise en compte de l'individu derrière la machine, dont on a expliqué que l'intérêt dépassait largement le champ de la mémorabilité des images, notre étude l'introduit dans le cadre de la prédiction de la mémorabilité des images. Un premier modèle *white box* a été proposé, qui révèle plusieurs facteurs individuels liés à la mémorabilité des images. Nous encourageons particulièrement de nouvelles études sur ce sujet, tant il nous semble qu'inclure l'idiosyncrasie dans les modèles de prédiction sera le principal amendement qu'on leur pourra apporter.

Dans la prochaine et dernière partie de cette thèse, nous nous intéressons aux modèles d'attention visuelle, dont le potentiel pour la prédiction de la mémorabilité d'images est intéressant. Nous présentons également le « film interactif émotionnel », un film dont le déroulement dépend des réponses émotionnelles du spectateur. Ce film constitue un outil intéressant pour étudier les liens entre attention visuelle et émotions, et potentiellement pour élargir nos travaux sur la mémorabilité aux vidéos.

IV

Oculométrie et modélisation de l'attention visuelle pour l'étude de l'émotion et de la mémorabilité

Introduction

Le regard est une porte ouverte vers les processus cognitifs. En témoignent la popularité croissante depuis plusieurs décennies de l'oculométrie pour étudier de tels processus (Just and Carpenter, 1976, Salvucci and Goldberg, 2000). Il est donc naturel que l'existence d'un lien entre l'attention visuelle et la mémorabilité ait été suggérée dès le commencement de son étude en vision par ordinateur (Isola et al., 2011a). Et les modèles computationnels d'attention visuelle, très répandus dans ce domaine de recherche, ont presque aussitôt été utilisés pour la prédiction de la mémorabilité d'images (Khosla et al., 2012b, Mancas and Le Meur, 2013, Celikkale et al., 2015). En 2013, Mancas et Le Meur ont réalisé une avancée intéressante en montrant que des caractéristiques liées à l'attention visuelle, extraites de l'image à l'aide de modèles d'attention visuelle, pouvaient avantageusement remplacer des caractéristiques de bas niveau pour la prédiction de la mémorabilité (Mancas and Le Meur, 2013). Selon ces auteur, le rôle de l'attention visuelle est important, et celle-ci devrait être davantage considérée, en lien avec les caractéristiques de bas et haut niveau des images, pour la prédiction de leur mémorabilité. Il est probable que l'engouement pour les modèles d'attention visuelle dans notre champ de recherche ne va pas s'essouffler d'ici longtemps.

Cependant, le comportement des modèles d'attention visuelle pour des images dont la coloration émotionnelle ou le degré de mémorabilité varie n'est pas bien connu. Dans le chapitre 11, nous étudions d'abord le lien entre attention visuelle, émotion et mémorabilité, à partir des données oculométriques enregistrées durant l'expérience décrite au chapitre 5. Ensuite, nous comparons la performance de modèles d'attention visuelle, et la croisons avec les scores d'émotion et de mémorabilité que nous avons précédemment obtenus pour 150 images.

Dans le chapitre 12, nous décrivons la mise au point du « film interactif émotionnel », un film dont le déroulement dépend des réactions émotionnelles du spectateur. L'émotion est mesurée en temps réel à l'aide d'un dispositif EEG qui fournit directement des données émotionnelles, en même temps qu'un oculomètre enregistre les mouvements oculaires du spectateur, ce qui permet de lier les variations émotionnelles à l'attention visuelle du spectateur et au contenu visionné. Comme nous l'expliquerons, cet outil possède un potentiel intéressant pour élargir nos travaux aux vidéos.

11

L'attention visuelle comme porte d'accès vers l'émotion et la mémorabilité des images

Après qu'*Isola et al.* ont exprimé l'intuition que la mémorabilité et l'attention visuelle pouvaient être liées (Isola et al., 2011b), Mancas et Le Meur ont montré que des caractéristiques de l'image liées à l'attention visuelle, extractibles à l'aide de modèles computationnels d'attention visuelle, pouvaient avantageusement remplacer certaines des caractéristiques de bas niveau des images utilisées par ces auteurs pour prédire la mémorabilité (Mancas and Le Meur, 2013). Ce chapitre porte, dans une première partie, sur les liens entre l'attention visuelle et la mémorabilité des images et l'émotion qu'elles véhiculent, à travers l'étude des fixations oculaires calculées à partir des données oculométriques collectées durant l'expérience décrite au chapitre 5. Nos résultats montrent des corrélations entre les scores d'émotion et de mémorabilité des images, et le nombre ainsi que la durée moyenne des fixations. Ils confirment l'intérêt de l'utilisation de caractéristiques de l'image liées à l'attention visuelle pour l'étude de la mémorabilité. En conséquence, nous nous intéressons, dans une deuxième partie, aux modèles d'attention visuelle, qui permettent, via la génération computationnelle de cartes de saillance, d'extraire des images des caractéristiques liées à l'attention visuelle, utilisables pour la prédiction automatique de la mémorabilité. Plus précisément, nous nous concentrons sur un problème précis : la performance de ces modèles est-elle liée à l'émotion véhiculée par les images, et à leur mémorabilité ? La réponse à ces questions n'a, à notre connaissance, jamais été apportée. Pour y répondre, nous comparons la performance de sept

modèles d'attention visuelle *bottom-up*, choisis pour leur performance et leur récence, en fonction des scores de mémorabilité et d'émotion que nous avons collectés pour 150 images. Les résultats montrent que la performance de certains modèles est corrélée aux scores d'émotion et de mémorabilité des images. Ces résultats pourront éclairer le choix d'un modèle d'attention visuelle plutôt qu'un autre pour l'étude de la mémorabilité ou de l'émotion véhiculée par les images. D'autre part, ils ouvrent la porte à l'amélioration des modèles d'attention visuelle en intégrant l'émotion véhiculée par les images et leur mémorabilité.

11.1 Introduction

Une scène visuelle naturelle est généralement complexe : elle contient de nombreux objets, qui ont des structures variées. Pour gérer cette surcharge d'information, l'évolution a pourvu le système visuel humain — dont les ressources sont limitées — de mécanismes attentionnels pour sélectionner les zones potentiellement intéressantes. Il y a donc une compétition entre les différentes informations pour leur traitement par le système visuel. Cette complexité des scènes visuelles naturelles se retrouve également, sous une forme qui se prête plus aisément à l'étude, dans les images.

De nombreux facteurs déterminent quelle partie d'une image sera sélectionnée ou exclue de la focalisation attentionnelle (Murray et al., 2011). Ces facteurs sont usuellement classés dans deux catégories : les facteurs *bottom-up* et les facteurs *top-down*, en raison des nombreuses preuves accumulées en faveur de l'existence d'un système à deux composantes, une composante *bottom-up* et une composante *top-down*, pour le contrôle du déploiement de l'attention visuelle dans une scène (p. ex. (Treisman and Gelade, 1980, Nakayama and Mackeben, 1989, Braun and Julesz, 1998)). Ainsi, en raison de la composante *bottom-up* du système visuel, le déploiement de l'attention dans une scène est influencée par les informations de la scène visuelle, telles qu'elles arrivent aux yeux. En particulier, la détection de la saillance est un processus non contrôlé : nos mécanismes *bottom-up* nous biaisent, de sorte que nous sélectionnons les stimuli sur la base de leur saillance. Selon Koch et Ullman (Koch and Ullman, 1987), la saillance d'une partie d'une scène est essentiellement déterminée par sa singularité en matière de couleur, d'orientation, de profondeur, etc., et les parties les plus saillantes forment de bons candidats pour les mécanismes de sélection attentionnels. Les mécanismes *bottom-up* ne déterminent cependant pas complètement la sélection attentionnelle. La répartition de l'attention visuelle dépend aussi de facteurs de plus haut niveau, les facteurs dits *top-down*, liés aux processus cognitifs tels que la connaissance de l'observateur, ses attentes, ses objectifs au moment de l'observation (Corbetta and Shulman, 2002). Typiquement, la tâche qu'il est demandé à un participant d'effectuer, ou le contenu sémantique d'une image, influencent la manière dont celui-ci va déployer son attention dans la scène visuelle (De Graef and Underwood, 2005, Tatler et al., 2011).

L'émotion que véhicule une scène visuelle est un des facteur importants qui vont influencer le déploiement de l'attention visuelle. Ainsi, il a été montré que l'exploration visuelle était sensible à la coloration et la signification émotionnelle de la scène (Christianson et al., 1991, Hermans et al., 1999). De plus, dans une scène visuelle, le fait pour un élément de se démarquer par sa coloration émotionnelle permet de sélectionner l'information rapidement et efficacement (p. ex. (Humphrey et al., 2012, Niu et al., 2012, Pourtois et al., 2013)). On conçoit d'ailleurs généralement que les réponses émotionnelles procèdent d'un système adaptatif (Cacioppo and Gardner, 1999). Dans une perspective évolutionniste, la réaction émotionnelle — par exemple, éviter un danger — est cruciale pour la survie ; aussi, l'évolution nous aura rendu sensibles en nous faisant associer une réponse émotionnelle aux informations importantes. Cela explique pourquoi les stimuli suscitant une émotion — en particulier une émotion de peur — sont traités préférentiellement aux stimuli neutres (p. ex. (Blanchette, 2006, Flykt, 2005, Öhman et al., 2001)). Comme pour les scène visuelles naturelles, les propriétés émotionnelles des images jouent également un rôle important dans l'exploration visuelle et la répartition de l'attention (Quirk and Strauss, 2001). Les propriétés émotionnelles d'une image peuvent aider à organiser son exploration visuelle, en dirigeant l'attention vers les régions les plus informatives (Christianson et al., 1991). De manière plus globale, les images suscitant de l'arousal sont plus largement explorées que les images neutres (Quirk and Strauss, 2001). Les images suscitant de l'émotion occasionneraient également un nombre de fixations plus important que les images neutres (Carniglia et al., 2012). Pour les raisons évoquées, l'émotion est un facteur important à prendre en compte pour les modèles d'attention visuelle.

En vision par ordinateur, les modèles d'attention visuelle sont principalement basés sur le concept de cartes de saillance (Riche et al., 2013). L'idée est que le regard humain tend à se diriger vers les régions qui se démarquent de leur contexte. Une carte de saillance est une carte bidimensionnelle qui fait correspondre à chaque partie d'une image (par exemple, chaque pixel) une probabilité d'attirer l'attention humaine (Itti and Koch, 2000). Ordinairement, la saillance à un endroit donné est déterminée en prenant en compte sa singularité par rapport à ce qui l'entoure en matière de couleur, d'orientation, de profondeur, etc. (Koch and Ullman, 1987). La probabilité de porter son regard sur une région particulière va également dépendre des mécanismes *top-down* ; cependant, en vision par ordinateur, les efforts pour modéliser l'attention visuelle humaine se sont surtout concentrés sur les mécanismes attentionnels *bottom-up* (Frintrop et al., 2010). Les modèles d'attention visuelle *bottom-up* déterminent la saillance à partir de propriétés de bas niveau de l'image : l'émotion qu'une image véhicule n'est pas prise en compte par ces modèles lors du calcul de la carte de saillance. Pourtant, la saillance émotionnelle crée une source de biais dans le traitement perceptuel, en conduisant l'attention vers les stimuli émotionnellement saillants (Simola et al., 2015) ; biais qui peut contrarier l'attraction de l'attention visuelle par les zones dont la saillance est définie

uniquement à partir des propriétés du stimulus (Niu et al., 2012). Les différents modèles *bottom-up* ont été développés en utilisant pour vérité terrain des bases d'images dont la charge émotionnelle n'a pas été évaluée ; et, à notre connaissance, leurs performances n'ont pas été évaluées pour des images émotionnellement chargées. Mettre en évidence un biais émotionnel — s'il existe — pourrait ouvrir la porte à l'amélioration de ces modèles en prenant en compte l'émotion véhiculée par les images.

Le déploiement de l'attention visuelle dans une image serait également lié à la mémorabilité des images (Mancas and Le Meur, 2013). Ces auteurs ont montré que la durée des fixations, qui reflèterait la profondeur du traitement de l'information visuelle (Henderson, 2007), baissait avec le degré de mémorabilité des images. Leur explication est que l'attention visuelle étant importante dans la mémorisation d'une image, la mémorabilité des images se traduit par un comportement visuel spécifique de l'observateur. Par suite de leurs résultats expérimentaux, les auteurs se sont intéressés aux modèles computationnels d'attention visuelle, pour extraire des images des informations liées à l'attention visuelle dans un objectif de prédiction computationnelle de la mémorabilité des images. Ils ont, dans cet objectif, utilisé plusieurs modèles d'attention visuelle (GBVS, RARE, etc.) pour générer des cartes de saillance, à partir desquelles ils ont calculé la couverture de l'image en matière de saillance, qui — montrent-ils — peut remplacer avantageusement certaines caractéristiques de bas niveau des images utilisées pour la prédiction par (Isola et al., 2011b). En particulier, ils obtiennent les meilleurs résultats en utilisant les cartes de saillance calculées par le modèle RARE (Riche et al., 2013). De la même manière que pour l'émotion véhiculée par les images, la performance des modèles d'attention visuelle n'a, à notre connaissance, pas été évaluée en fonction de la mémorabilité des images.

Dans la suite de ce chapitre, nous utilisons, dans une première partie, les données d'oculométrie collectées dans notre étude, en plus des scores d'émotion et de mémorabilité des images cibles. En particulier, de ces données nous extrayons le nombre et la durée des fixations occasionnées par les différentes images cibles lors des deux tâches de reconnaissance que nos 50 participants ont passées, et testons la corrélation de ces facteurs avec les scores de mémorabilité, d'arousal et de valence des images. Compte tenu des résultats obtenus par (Mancas and Le Meur, 2013), nous nous attendons à ce que la durée moyenne des fixations soit d'autant plus longue que les images qui les ont occasionnées sont mémorables. Étant donnés les résultats obtenus par (Carniglia et al., 2012), nous nous attendons également à ce que les images les plus chargées émotionnellement occasionnent un nombre de fixations moyen plus important ainsi qu'une durée moyenne de fixation plus longue que les images plus neutres.

Dans un second temps, nous évaluons la performance de sept modèles d'attention visuelle différents en fonction des scores d'émotion et de mémorabilité de nos images cibles. Si la première partie de nos analyses corrobore l'hypothèse d'un lien entre émotion et mémorabilité et attention visuelle, alors nous attendrons de nos ana-

lyses qu'elles révèlent d'éventuels biais de performance dans les modèles, avec des modèles meilleurs/moins bons pour certains degrés d'émotion ou/et de mémorabilité des images.

11.2 Analyses des données oculométriques

Un oculomètre enregistrait les mouvements des yeux des participants à notre étude pendant qu'ils réalisaient les différentes tâches de l'expérience (i.e. les deux tests de mémoire et la tâche de notation des images sur les dimensions d'arousal et de valence). Les analyses présentées dans la suite de ce chapitre portent sur les données oculométriques de 49 participants (sur les 50 que comptait notre expérience, le sujet 47 n'ayant pas passé la dernière partie de l'expérience) ; elles se concentrent sur le nombre et la durée moyens des fixations pour les 150 images cibles, dans chacun des deux tests de mémoire et la tâche de notation.

11.2.1 Fixations et scores des images de notre base

	Mémorabilité	Nbre fix	Durée fix	Arousal	Valence
Mémorabilité test 1	1				
Nbre fix test 1	$.23^{**}$	1			
Durée fix test 1	$-.10$	$-.42^{***}$	1		
Arousal	$.23^{**}$	$.23^{**}$	$-.19^{*}$	1	
Valence	$-.27^{***}$	$-.09$	$.11$	$-.52^{***}$	1
Mémorabilité test 2	1				
Nbre fix test 2	$.11$	1			
Durée fix test 2	$-.14$	$-.40^{***}$	1		
Arousal	$.42^{***}$	$.20^{*}$	$-.17^{*}$	1	
Valence	$-.28^{***}$	$-.12$	$.12$	$-.52^{***}$	1

TABLE 11.1 – Corrélations entre les scores de mémorabilité, d'arousal et de valence des images de notre base de données et le nombre et la durée des fixations sur les $2s$ de visionnage de l'image, pour le premier test de mémoire (en haut) et le second test (en bas). Les scores d'émotion ayant été attribués dans une troisième phase de notation, ils sont les mêmes pour les deux tests. $^{*}p < .05$. $^{**}p < .01$

La table 11.1 présente les corrélations entre le nombre et la durée moyens des fixations occasionnées par les 150 images cibles lors de leur première occurrence, soit pour une durée de $2s$, et leurs scores de mémorabilité, d'arousal et de valence. Nous n'avons

pas pris en compte les données oculométriques enregistrées durant le visionnage d'une image alors qu'elle était répétée, pour la raison que la réponse des participants pouvait abréger la présentation de l'image, vue, dans ce cas, moins de $2s$. On peut remarquer que le pattern des résultats est similaire pour les deux tests de mémoire : les corrélations significatives pour un test de mémoire le sont pour l'autre, et les tendances vont dans le même sens. Une exception concerne la mémorabilité, dont la corrélation avec le nombre de fixations n'est pas significative pour le second test de mémoire alors qu'elle l'est pour le premier test.

Concernant la mémorabilité des images, nos résultats montrent une corrélation linéaire positive significative entre le nombre moyen de fixations et la mémorabilité des images pour le premier test de mémoire ($r = .23, p < .01$), qui n'est pas confirmée par le second test de mémoire ($r = .11, p = .168$), quoique la tendance de la corrélation soit similaire. Ce résultat confirme un lien entre mémoire et attention visuelle, et rend d'autant plus intéressant l'utilisation de modèles d'attention visuelle dans le cadre de la prédiction de la mémorabilité des images.

En revanche, nous ne trouvons pas de corrélation significative entre la durée moyenne de fixation et la mémorabilité des images, ni pour le premier test de mémoire ($r = -.10, p = .222$), ni pour le second test ($r = -.14; p = .078$). Les tendance vont à l'inverse des résultats obtenus par (Mancas and Le Meur, 2013), qui ont trouvé que la durée moyenne des fixations était d'autant plus longue que les images qui les avaient occasionnées étaient mémorables. Sur ce point, il faut noter la corrélation négative significative entre le nombre moyen de fixations et la durée moyenne des fixations, observée pour le premier test de mémoire ($r = -.42, p < .001$) et pour le second test ($r = -.40, p < .001$). Cette corrélation négative peut s'expliquer par le fait que la tâche était à temps contraint, comme dans l'expérience ayant servi à créer la base de (Isola et al., 2011b). Le fait que nos participants savaient qu'ils n'avaient que $2s$ pour mémoriser chaque image a pu inverser la tendance par rapport à (Mancas and Le Meur, 2013), dont les participants voyaient les images $5s$.

Les résultats obtenus pour l'arousal (illustrés par la figure 11.1, à gauche) sont particulièrement intéressants : l'arousal est à la fois positivement corrélé au nombre moyen de fixations pour le premier test de mémoire ($r = .23, p < .01$) et pour le second test de mémoire ($r = .20, p < .05$), et négativement corrélé à la durée moyenne des fixations pour le premier test ($r = -.19, p < .05$) et le second test ($r = -.17, p < .05$). Ces résultats suggèrent que lorsque l'arousal suscité par une image augmente, les individus tendent à fixer leur attention sur un plus grand nombre de zones, et à passer moins de temps sur les différentes zones d'intérêt.

Nous n'avons pas trouvé de corrélation linéaire significative entre la valence et le nombre ou la durée moyens des fixations. Sur la figure 11.1 (à droite), une forme en U semblent cependant se dessiner — qui pourrait s'expliquer par la relation, de cette forme, qu'entretiennent l'arousal et la valence. Pour cette raison, nous avons séparé

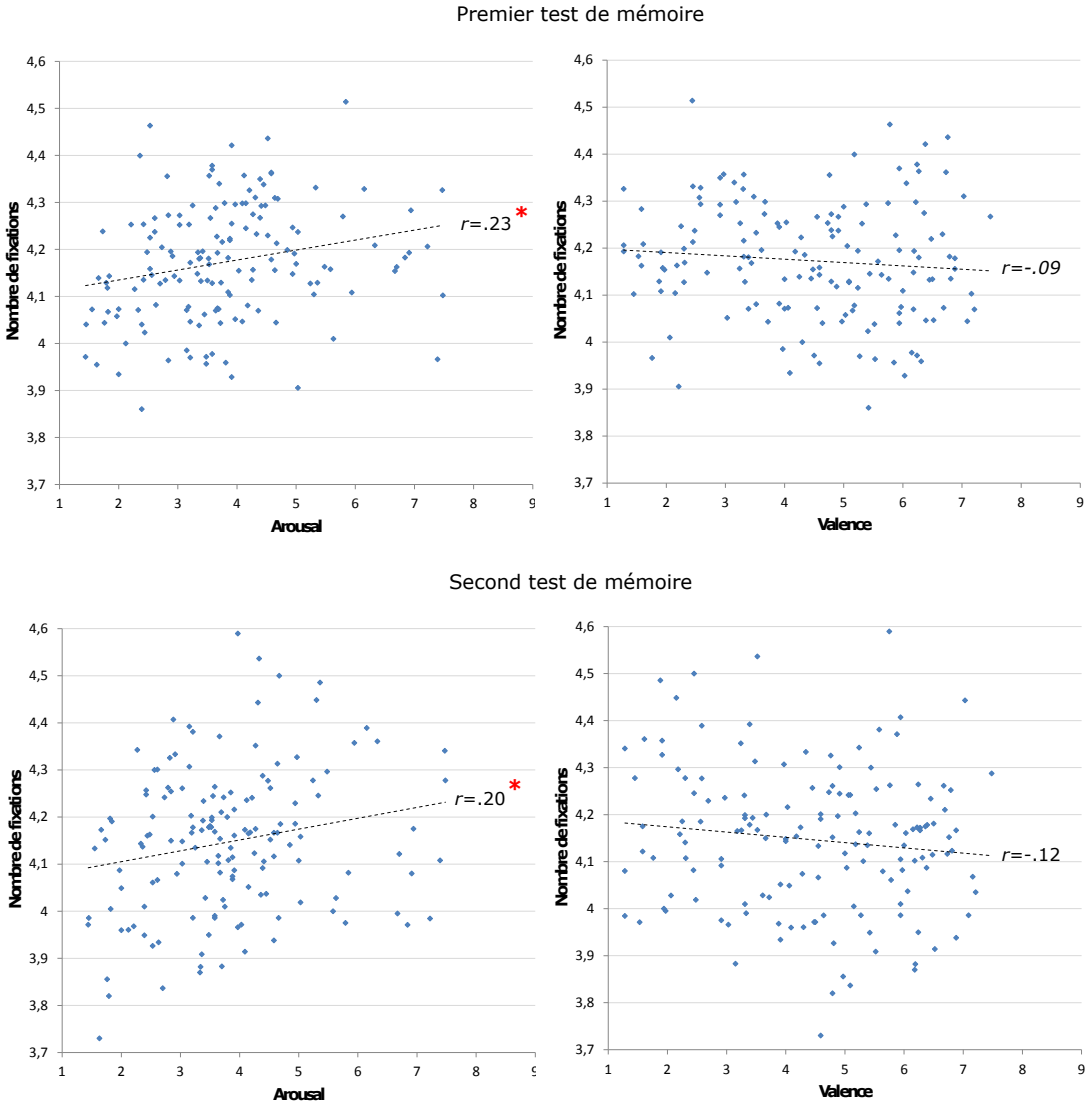


FIGURE 11.1 – Nuage de points des 150 images cibles de notre étude, selon leurs scores d'arousal (à gauche) ou de valence (à droite) et le nombre moyen de fixations qu'elles ont occasionnées lors du premier test de mémoire (en haut) et du second test de mémoire (en bas). Les lignes pointillées représentent la tendance linéaire. Une * signifie que la corrélation est significative.

les images en deux groupes selon leur valence positive (> 5) ou négative (< 5). Pour les images de valence négative, nous n'avons pas observé de corrélation linéaire significative entre la valence et le nombre de fixations pour le premier test de mémoire ($r = -.12, p = .274$), mais nous avons observé une corrélation négative significative pour le second test ($r = -.22, p < .01$). De plus, nous n'avons pas observé de corrélation significative entre la valence et la durée moyenne de fixation, ni pour le premier test de mémoire ($r = .14, p = .185$), ni pour le second ($r = .19, p = .079$). Pour les images de valence positive, nous n'avons pas observé de corrélation linéaire significative entre la valence et le nombre de fixations ni pour le premier test de mémoire ($r = .16, p = .228$), ni pour le second test ($r = .01, p = .958$). Nous avons cependant observé une corrélation négative significative entre la valence et la durée moyenne de fixation pour le premier test de mémoire ($r = -.26, p < .05$), mais pas pour le second test ($r = -.22, p = .081$). De ces résultats, qui vont dans le même sens pour les deux tests de mémoire, une tendance semble se dessiner : plus la valence d'une image est négative ou positive, plutôt que neutre, plus le nombre de fixations tend à être élevé, et la durée des fixations basse.

On peut remarquer dans la figure 11.1 que les fixations sont relativement courtes. Il est probable que cela s'explique par la particularité de la tâche. En effet, comme nous l'avons évoqué, la tâche que réalise un participant est susceptible d'influencer la manière dont il déploie son attention dans une scène visuelle (De Graef and Underwood, 2005). Les participants à notre étude savaient qu'ils avaient peu de temps (i.e. $2s$) pour mémoriser et reconnaître les images, ce qui les aura certainement incités à ne pas s'attarder sur les différentes parties de l'image.

11.3 Performance des modèles de saillance

Les résultats présentés dans la section précédente montrent des associations entre, d'un côté, les émotions véhiculées par les images et leur mémorabilité, et de l'autre le nombre, ainsi que la durée moyenne, des fixations. Ils corroborent l'hypothèse d'un lien entre attention visuelle et émotion et mémorabilité. Dans cette section, nous cherchons à répondre à la question suivante : puisque l'attention visuelle est liée à l'émotion véhiculée par les images et à leur mémorabilité, la performance des modèles computationnels d'attention visuelle est-elle également dépendante de ces facteurs ?

11.3.1 Méthode

Notre analyse de la performance de modèles d'attention visuelle a porté sur les 150 images pour lesquelles nous avons collecté des scores d'arousal, de valence et de mémorabilité dans notre étude. Pour réaliser cette analyse, nous avons procédé par étapes : d'abord (1) nous avons calculé des cartes de saillance pour chaque image à partir de

notre vérité terrain ; ensuite (2), nous avons généré des cartes de saillance à partir de sept modèles d'attention visuelle différents ; puis (3) nous avons calculé la performance des modèles comme l'écart entre les cartes de saillance *vérité terrain*[1] et les cartes de saillance générées par les modèles ; enfin (4), nous avons comparé la performance des modèles aux scores d'arousal, de valence et de mémorabilité des images.

11.3.2 Calcul des cartes de densité de fixation

Nous avons, dans un premier temps, calculé des cartes de densité de fixation à partir des fixations visuelles des participants à notre étude pour les 150 images cibles, en suivant la méthode décrite par (Le Meur and Baccino, 2013). Les cartes de densité de fixation ont été calculées séparément pour un temps de visionnage de deux secondes, correspondant à la durée de présentation d'une image lors de sa première occurrence durant les tests de mémoire, et pour un temps de visionnage de six secondes, correspondant à la durée de présentation d'une image durant la tâche de notation des images cibles sur les dimensions d'arousal et de valence. On notera qu'outre la durée de visionnage, le type de tâche demandé change également, ce qui — comme nous l'avons évoqué — peut influencer le déploiement de l'attention visuelle des participants. Dans les résultats présentés dans la suite de ce chapitre, les cartes de densité de fixation correspondant aux deux tests de mémoire (très similaires) ont été fusionnées. Un exemple de ces cartes de densité de fixation est proposé dans la figure 11.2 (les cartes de densité de fixation du premier et du second test de mémoire n'ont pas encore été fusionnées).

11.3.3 Modèles évalués

Dans un second temps, nous avons sélectionné des modèles d'attention visuelle. La génération computationnelle de cartes de saillance est un problème ouvert dont l'intérêt est croissant en vision par ordinateur (Murray et al., 2011), avec un nombre exponentiel d'articles proposant des algorithmes de calcul de cartes de saillance publiés ces dernières années (Riche et al., 2013). Pour notre étude, nous avons sélectionné sept modèles *bottom-up*, en prenant en compte leur récence mais également leur performance, évaluée par (Bylinskii et al., 2016) : le modèle d'Itti et Koch (Itti et al., 1998), GBVS (pour *Graph-Based Visual Saliency*) (Harel et al., 2006), Self-Resemblance Saliency (Seo and Milanfar, 2009), Saliency estimation (Murray et al., 2011), Fast and efficient saliency (Tavakoli et al., 2011), RARE (Riche et al., 2013) et CovSal (Erdem and Erdem, 2013). Nous avons généré pour chacune des 150 images sept cartes de saillance, correspondant à ces sept modèles ; un exemple est donné dans la figure 11.2. Pour une

[1] Dans la suite de ce chapitre, nous appelons cartes de densité de fixation ces cartes de saillance issues de la vérité terrain, pour bien les distinguer des cartes de saillance générées par les modèles d'attention visuelle, que nous appelons simplement cartes de saillance.

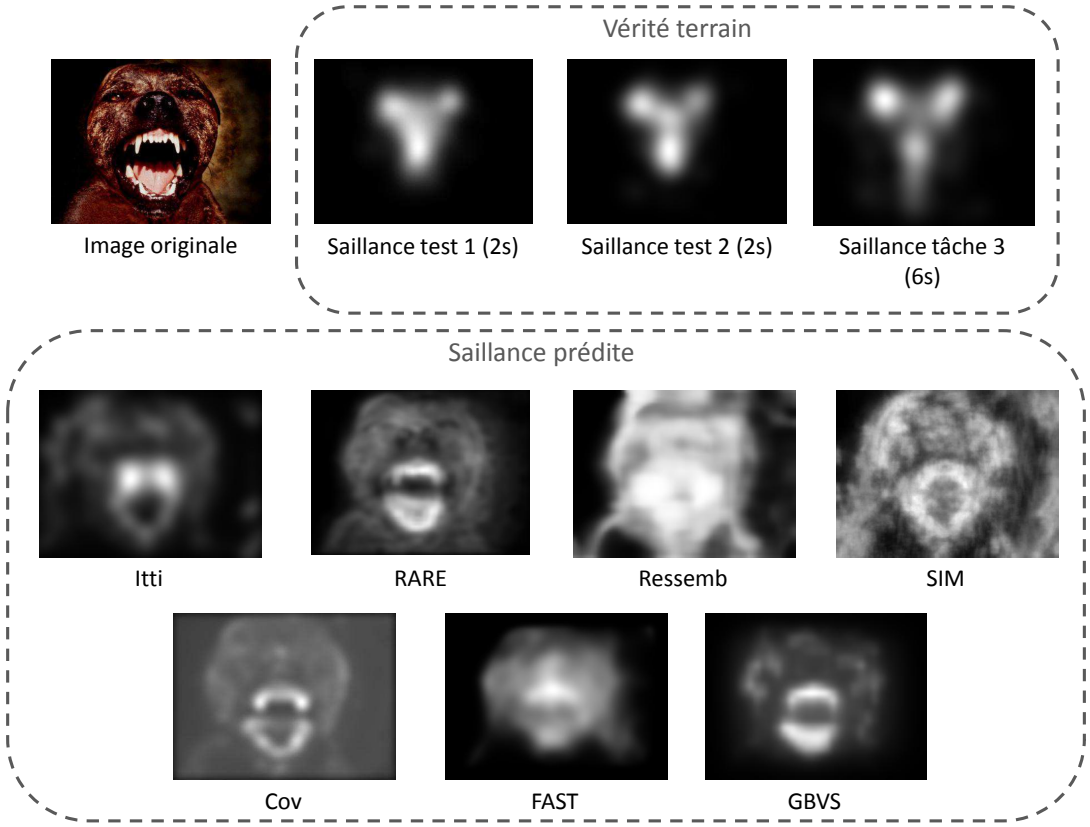


FIGURE 11.2 – L'image 1300 de l'IAPS et les cartes de saillance correspondantes. En haut, les cartes de densité de fixation ont été calculées à partir de nos données oculométriques, collectées durant les différentes tâches effectuées par les participants (les deux tests de mémoire, où les participants voyaient une image durant $2s$, et la tâche de notation des images sur les dimensions d'arousal et de valence, où les images étaient vues $6s$). En bas, les cartes générées par les différents modèles d'attention visuelle sélectionnés.

raison de lisibilité, dans cette figure et dans la suite de ce chapitre, nous ferons référence aux différents modèles par les noms suivants : « itti »pour (Ito et al., 1998), « gbvs »pour (Harel et al., 2006), « resemb »pour (Seo and Milanfar, 2009), « sim »pour (Murray et al., 2011), « fast »pour (Tavakoli et al., 2011), « rare »pour (Riche et al., 2013) et « cov »pour (Erdem and Erdem, 2013).

11.3.4 Mesures utilisées pour déterminer la performance des modèles

Pour déterminer la performance des modèles d'attention visuelle, nous avons comparé chaque carte de saillance à la carte de densité de fixation qui lui correspondait. Nous avons utilisé quatre méthodes de mesure, complémentaires, couramment utilisées pour évaluer le degré de similarité entre deux cartes de saillance : deux mesure basées sur la corrélation entre les cartes de saillance (corrélations de Pearson et de Spearman), la divergence de Kullback-Leibler et l'aire sous la courbe ROC (AUC). Ces différentes méthodes de mesure, que nous décrivons brièvement ici, ont été décrites en détail dans (Le Meur and Baccino, 2013). Chacune de ces méthodes nous donnera une indication de la performance des sept modèles d'attention visuelle évalués pour prédire où des observateurs réels vont regarder.

Le coefficient de corrélation de Pearson r entre deux cartes de saillance — dans notre cas une carte générée par un modèle et une carte de densité de fixation — est utilisé pour déterminer si les données partagent la même structure (p. ex. dans (Le Meur et al., 2006, Jost et al., 2005)). Nous l'avons calculé pour chaque paire possible carte de densité de fixation-carte de saillance ; nous avons également calculé les coefficient de corrélation de rang de Spearman, parfois utilisés dans ce même objectif (Toet, 2011).

La divergence de Kullback-Leibler (KLD) est utilisée pour estimer la dissimilarité globale entre deux fonctions de densité de probabilité. Cette divergence n'est pas une distance. D'autre part, elle n'est pas symétrique : $KLD(S1, S2) \neq KLD(S2, S1)$ ($S1$ et $S2$ renvoyant ici à une carte de densité et une carte de saillance, respectivement). Pour cette raison, nous avons calculé la moyenne de $KLD(S1, S2)$ et de $KLD(S2, S1)$. Une valeur de zéro indique que les deux fonctions de densité sont strictement égales ; plus la divergence est grande, plus les fonctions sont dissemblables. KLD n'a pas de limite supérieure bien définie, ce qui constitue, selon (Le Meur and Baccino, 2013), une importante faiblesse de cette méthode. Il faut noter que, contrairement aux mesures de corrélation et à l'AUC, plus la KLD augmente, plus la dissemblance entre les cartes de saillance est forte.

L'analyse ROC (Green and Swets, 1966), enfin, est probablement la méthode la plus populaire pour évaluer le degré de similarité de deux cartes de saillance (Le Meur and Baccino, 2013). Pour cette raison, et pour une question de lisibilité, certains graphiques présentés par la suite porteront uniquement sur cette mesure. Ordinairement, l'analyse

ROC implique deux jeux de données : le premier correspondant à la vérité terrain, et le second aux prédictions. La mesure de l'aire sous la courbe ROC (AUC) indique une performance globale de classification : une valeur de 1 indique une classification parfaite ; le niveau de hasard est 0.5.

11.3.5 Résultats

L'analyse des résultats comprend trois parties : dans une première partie, nous comparons la performance moyenne des modèles sur l'ensemble des images ; ensuite, nous calculons les corrélations entre la performance des modèles et les scores d'émotion et de mémorabilité des images ; enfin, nous affinons notre analyse en créant des groupes d'images sur la base de leurs scores d'émotion pour donner une meilleure idée du comportement des modèles selon l'émotion véhiculée par les images.

Performance globale des modèles d'attention visuelle

La performance moyenne des modèles d'attention visuelle sur les 150 images est représentée dans la figure 11.3, pour chacune des quatre mesures utilisées. Pour l'AUC, l'axe des ordonnées commence à 0.5, qui correspond au niveau du hasard. Concernant la KLD, contrairement aux autres mesures, plus sa valeur est élevée, plus la divergence moyenne entre les cartes de saillance et les cartes de densité de fixation est importante ; il faut également rappeler que la KLD n'est pas une distance.

En observant la figure 11.3 dans sa globalité, on remarque que, quelle que soit la mesure utilisée, la performance relative des modèles est globalement la même. Les modèles les plus/moins performants selon une des mesures utilisées sont également les modèles plus/moins performants pour les autres mesures, et ceci dans les deux conditions de visionnage (i.e. $2s$ en tâche de reconnaissance et $6s$ en tâche de notation).

Plus le temps de visionnage de l'image est long, plus le nombre d'informations *top-down* qui participent au déploiement de l'attention visuelle tend à augmenter. Les sept modèles sélectionnés étant *bottom-up*, on pourrait s'attendre à ce que leur performance soit meilleure pour un visionnage de deux secondes que de six secondes. Si l'on s'intéresse à la mesure de référence, l'AUC, un test de Kruskal-Wallis ($\chi^2(1) = 8.04, p < .005$) le confirme : les modèles (pris ensemble) sont plus performants pour prédire le déploiement de l'attention humaine pour un temps de visionnage de deux secondes ($\mu = 0.832; med = 0.869$) que de six secondes ($\mu = 0.824; med = 0.853$). Cependant, cela cache des disparités : l'effet est tiré uniquement par les modèles *cov* et *fast*, tandis que pour les autres modèles il n'y a pas de différence significatives entre les deux conditions de visionnage. Les autres mesures semblent confirmer les résultats observés pour *cov* et *fast* ; on observe cependant qu'un pattern inverse tend à apparaître pour les autres modèles lorsqu'on considère les mesures de corrélation. La méthode de mesure utilisée pour estimer la performance des modèles de prédiction est donc importante.

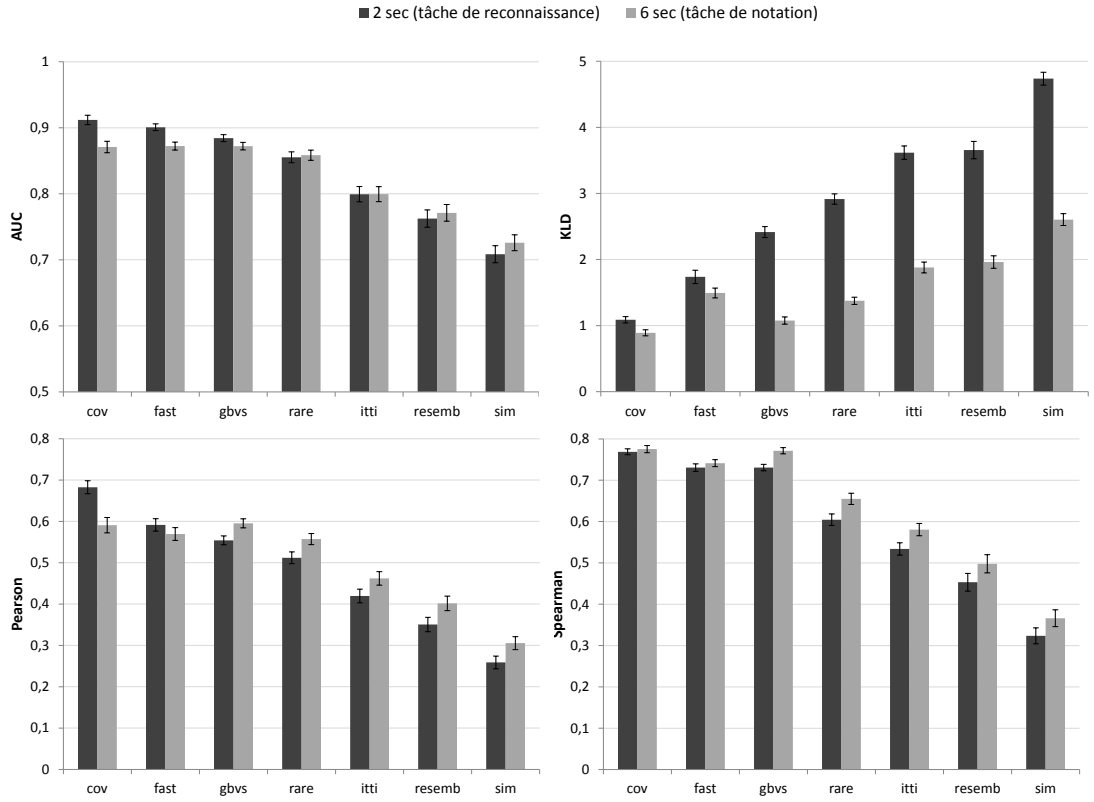


FIGURE 11.3 – Performance globale (sur les 150 images) des sept modèles d'attention visuelle. Les quatre mesures utilisées comparent les cartes de saillance aux cartes de densité de fixation. Les barres d'erreur correspondent aux erreurs-types des moyennes.

Corrélations entre la performance et l'arousal

Dans l'objectif d'évaluer la performance des modèles d'attention visuelle pour des images véhiculant des émotions différentes, nous avons calculé les coefficients de corrélation entre la performance des modèles telle que mesurée par les quatre méthodes de mesure utilisées et les scores d'arousal des images de notre base de données. Les résultats sont présentés dans la figure 11.2.

	gbvs	**itti**	**rare**	**cov**	**sim**	**fast**	**resemb**
Tests de mémoire ($2sec$)							
Pearson	.06	.16	$.20^{*}$	$-.29^{***}$	$.21^{**}$	.07	$.19^{*}$
Spearman	-.05	.08	.09	$-.27^{***}$	$.19^{*}$	.04	$.18^{*}$
KLD	-.03	-.13	-.09	$.23^{**}$	$-.20^{*}$	-.05	$-.28^{***}$
AUC	.00	.12	$.20^{*}$	$-.34^{***}$	$.19^{*}$	-.03	.14
Tâche de notation ($6sec$)							
Pearson	-.07	.11	.10	$-.23^{**}$	$.21^{**}$	.03	$.20^{*}$
Spearman	-.09	.09	.07	$-.26^{***}$	$.20^{*}$	-.03	$.21^{*}$
KLD	.09	.00	.07	$.28^{***}$	-.02	.10	$-.20^{*}$
AUC	-.07	.10	.13	$-.28^{***}$	$.18^{*}$	-.02	.14

TABLE 11.2 – Corrélations entre la performance des modèles et les scores d'arousal des 150 images cibles. En haut, les cartes de densité de fixation ont été calculées pour une durée de visionnage de deux secondes (durant les tests de mémoire) ; en bas, pour une durée de six secondes (durant la tâche de notation des images sur les dimensions d'arousal et de valence). $^{*} < .05$; $^{**}p < .01$; $^{***}p < .001$.

Ces résultats sont particulièrement intéressants pour *cov*, *sim* et *resemb* : ces trois modèles montrent, pour les deux conditions de visionnage (deux secondes et six secondes), et pour (presque) toutes les mesures, des performances corrélées aux scores d'arousal des images. Cela suggère qu'il y a un biais émotionnel dans certains des modèles testés. Concernant *cov*, sa performance est négativement corrélée au degré d'arousal des images, ce qui signifie que la performance du modèle pour prédire le déploiement de l'attention visuelle humaine dans des images tend à baisser lorsque le degré d'arousal suscité par les images augmente. On observe le pattern inverse pour *sim* et *resemb*, qui tendent, au contraire, à être plus performants lorsque l'arousal véhiculé par les images augmente. A noter que, à moindre mesure, on observe une tendance similaire à ces deux modèles pour *rare*. Il y a donc un biais dans certains des modèles, qui ne va

pas nécessairement dans le même sens.

Corrélations entre la performance et les scores de mémorabilité des images

Afin d'évaluer la performance des modèles d'attention visuelle pour des images plus ou moins mémorables, nous avons calculé les coefficients de corrélation entre la performance des modèles et les scores de mémorabilité des images. Les résultats sont présentés dans les figures 11.3 et 11.4, qui portent respectivement sur les scores de mémorabilité obtenus grâce au premier test de mémoire (la récupération mnésique avait lieu quelques minutes après la phase d'encodage) et ceux obtenus grâce au second test de mémoire (la récupération avait lieu un jour après l'encodage).

	gbvs	itti	rare	cov	sim	fast	resemb
Tests de mémoire ($2sec$)							
Pearson	-0.01	-0.14	0.02	-0.08	-0.06	-0.16^{*}	0.09
Spearman	-0.11	-0.11	-0.04	-0.12	-0.03	-0.12	0.05
KLD	0.08	0.10	0.10	0.13	0.08	0.10	0.01
AUC	-0.07	-0.11	0.01	-0.12	-0.06	-0.10	0.06
Tâche de notation ($6sec$)							
Pearson	-0.07	-0.17^{*}	-0.02	-0.05	-0.08	-0.14	0.06
Spearman	-0.17^{*}	-0.14	-0.01	-0.15	-0.03	-0.19^{*}	0.05
KLD	0.12	0.11	0.06	0.13	0.06	0.12	0.00
AUC	-0.13	-0.14	-0.01	-0.11	-0.09	-0.06	0.05

TABLE 11.3 – Corrélations entre la performance des modèles et les scores de mémorabilité des 150 images cibles calculés à partir des résultats au premier test de mémoire (passé quelques minutes après la phase d'encodage mnésique). $^{*} < .05$; $^{**}p < .01$.

Pour les scores de mémorabilité obtenus grâce au premier test de mémoire (figure 11.3), les tendances sont cohérentes entre les différentes mesures, et entre les conditions de visionnage (deux secondes *vs.* six secondes). Cependant, peu de corrélations apparaissent statistiquement significatives, et pour celles qui l'apparaissent, elles ne sont pas confirmées pour les autres mesures. Pour les scores de mémorabilité obtenus grâce au second test de mémoire, en revanche, nous pouvons observer dans la table 11.4, pour *cov* et *fast*, des corrélations significatives entre la performance des modèles et les scores de

	gbvs	itti	rare	cov	sim	fast	resemb
Tests de mémoire ($2sec$)							
Pearson	-0.02	-0.06	0.09	-0.25^{**}	0.02	-0.22^{**}	0.15
Spearman	-0.12	-0.08	0.06	-0.19^{*}	0.06	-0.18^{*}	0.11
KLD	0.13	0.13	0.11	0.25^{**}	0.07	0.25^{**}	0.03
AUC	-0.08	-0.06	0.08	-0.24^{**}	0.02	-0.21^{*}	0.10
Tâche de notation ($6sec$)							
Pearson	-0.11	-0.14	0.03	-0.19^{*}	-0.02	-0.22^{**}	0.12
Spearman	-0.09	-0.08	0.11	-0.15	0.06	-0.19^{*}	0.12
KLD	0.15	0.17^{*}	0.11	0.18^{*}	0.10	0.19^{*}	0.06
AUC	-0.11	-0.09	0.06	-0.18^{*}	-0.02	-0.13	0.08

TABLE 11.4 – Corrélations entre la performance des modèles et les scores de mémorabilité des 150 images cibles calculés à partir des résultats au second test de mémoire (passé un jour après la phase d'encodage mnésique). $^{*} < .05$; $^{**}p < .01$.

mémorabilité des images, qui sont corroborées par les différentes mesures et cohérentes entre les deux conditions de visionnage.

Les résultats montrent que les modèles *cov* et *fast* tendent à être moins performants lorsque la mémorabilité des images augmente. Alors que les performances de ces deux modèles étaient, respectivement, négativement et positivement corrélées avec les score d'arousal des images, leurs performances sont toutes deux négativement corrélées à la mémorabilité des images. Cela suggère que l'arousal, dont on a montré qu'il était corrélé à la mémorabilité, n'est pas (entièrement) responsable des différences de performance en fonction du degré de mémorabilité des images.

On peut également noter, en comparant les deux tables, que les tendances sont similaires pour les scores de mémorabilité obtenus grâce au premier et au second test de mémoire ; mais que les corrélations sont (dans quasiment tous les cas) plus fortes pour les scores obtenus au second test que ceux obtenus au premier.

11.3.6 Performances locales des modèles

Le coefficient de corrélation linéaire nous donne un aperçu de la performance globale d'un modèle en fonction de l'arousal véhiculé par les images. Pour pouvoir mieux apprécier visuellement le comportement local des modèles pour des images suscitant soit un arousal faible, soit modéré, soit élevé, nous avons découpé les images en trois

groupes de taille équivalente sur la base de leurs scores d'arousal : le premier groupe comprend les 50 images dont le score de mémorabilité est le plus faible, et ainsi de suite. Ensuite, nous avons mesuré (avec la méthode AUC) la performance moyenne des modèles pour ces trois groupes d'images. Les résultats sont présentés dans la figure 11.4.

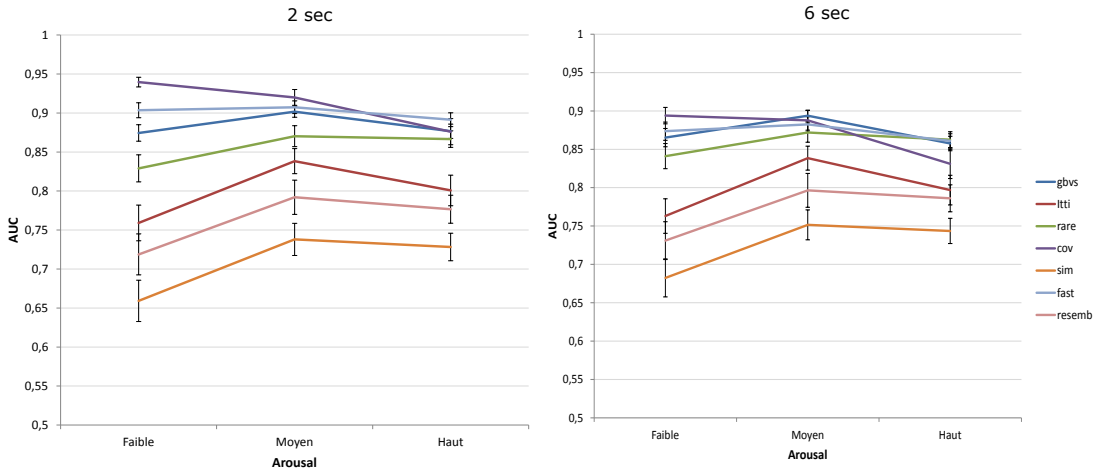


FIGURE 11.4 – Performance moyenne des modèles en fonction des trois groupes d'images (faible, moyen et haut arousal). Les cartes de densité de fixation ont été calculées pour une durée de visionnage de 2 secondes (à gauche) et de 6 secondes (à droite). Les barres d'erreur correspondent aux erreurs-types des moyennes.

Les dimensions d'arousal et de valence étant liées (voir la section 6.2 du chapitre 6), nous avons également mesuré la performance moyenne des modèles pour les trois clusters *arousal-valence* présentés dans la figure 8.2 du chapitre 8. Les trois clusters séparent grossièrement les images négatives, qui, suscitent un arousal fort, des images neutres, qui suscitent pas ou peu d'arousal, et des images positives, qui suscitent un arousal modéré. Les résultats sont présentés dans la figure 11.5.

Les graphiques 11.4 et 11.5 apportent des précisions par rapport aux tables de corrélations ; en particulier, si l'on s'intéresse à un certain types d'images correspondant aux groupes présentés dans ces graphiques, on pourra s'aider de ces derniers pour le choix du modèle d'attention visuelle le plus performant.

11.4 Discussion

Dans ce chapitre, nous avons trouvé, en étudiant le nombre et la durée des fixations, que l'attention visuelle était liée à l'émotion véhiculée par les images et à leur mémorabilité. Nous avons également évalué la performance de différents modèles computationnels

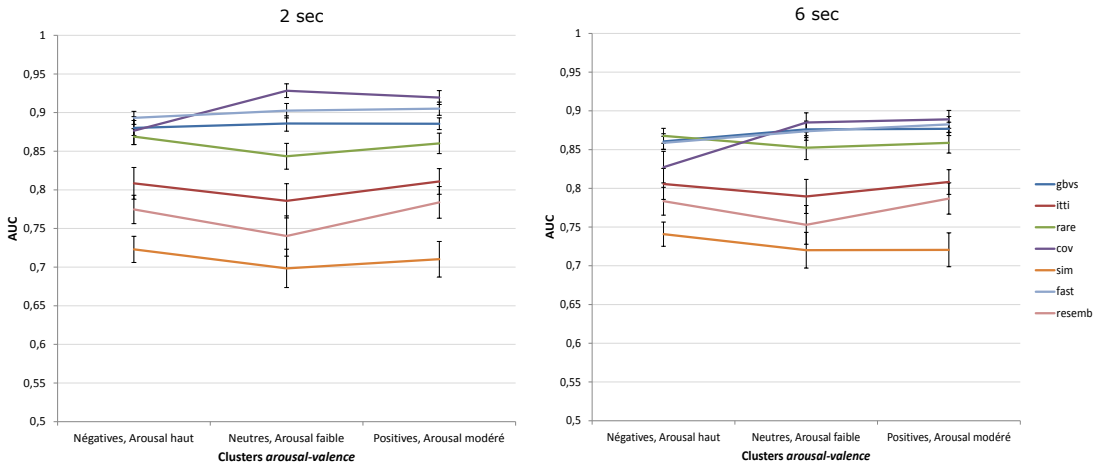


FIGURE 11.5 – Performance moyenne des modèles en fonction des trois clusters *arousal-valence* créés à l'aide d'un algorithme de partitionnement en k-moyennes ($k = 3$). Les barres d'erreur correspondent aux erreurs-types des moyennes.

d'attention visuelle, et montré notamment que la performance de certains d'entre eux pour prédire où, dans une image, des observateurs humains vont regarder, était liée à l'arousal suscité par les image et à leur degré de mémorabilité. Ces résultats révèlent donc des biais dans les modèles, que nous pourrions imaginer corriger en prenant en compte la mémorabilité des images et l'émotion qu'elles véhiculent, prédites par des modèles de prédiction du type de celui de (Isola et al., 2011b) ou de MemoNet.

Sur le comportement oculaire des participants

Nous avons trouvé une corrélation positive entre le nombre moyen de fixations oculaires et la mémorabilité des images plus forte pour le premier test de mémoire que pour le second. On peut exclure un effet de la rétention mnésique, puisque nous avons pris en compte dans nos analyses uniquement les données oculométriques durant le visionnage des images lors de leur première occurrence, et pas lors de leur répétition. (Suivant notre plan d'expérience, lorsque cette première occurrence advenait au second test de mémoire, l'image était pour le participant une image de remplissage ; pour d'autres participants, elle était une image cible.) Reste les différences entre les deux tests de mémoire. Parmi celles-ci, il nous semble important d'en citer deux : d'une part, la tâche était un peu différente au second test de mémoire, puisque les participants n'avaient plus à mémoriser de nouvelles images mais juste à en reconnaître ; d'autre part, la tâche de mémoire était plus difficile au second test qu'au premier, puisque les participants devaient reconnaître des images encodées un jour auparavant, contre quelques minutes auparavant pour le premier test. Or, comme nous l'avons dit, la tâche effectuée par un

individu est susceptible d'influencer le déploiement de son attention visuelle. On peut, dès lors, penser que ces différences entre les tâches ont influencé le déploiement de l'attention visuelle de nos participants.

Nous n'avons pas trouvé de corrélation significative entre la durée moyenne des fixations et la mémorabilité des images. Les tendances observées vont à l'inverse des résultats obtenus par (Mancas and Le Meur, 2013), qui ont trouvé que la durée moyenne des fixations était d'autant plus longue que les images qui les avaient occasionnées étaient mémorables. Comme explication plausible, nous avons invoqué la différence de conditions de visionnage dans nos études. On peut ajouter que les effets montrés par ces auteurs sont ténus : pour observer une différence significative, ils ont redécoupé arbitrairement l'ensemble de leurs images, constituant avec certaines d'entre elles deux groupes, l'un contenant les 20 images les plus mémorables, et l'autre les 20 images les moins mémorables. D'autre part, nos résultats ne sont que des tendances. En conclusion, de nouvelles études sont nécessaires afin de confirmer ou infirmer l'un ou l'autre de ces résultats.

Nous avons trouvé que plus le score d'arousal d'une image augmentait, plus le nombre de fixations tendait à augmenter. De la même manière, nous avons trouvé, pour les images négatives, que plus le score de valence tendait à augmenter, plus le nombre de fixations baissait ; et, pour les images positives, que plus le score de valence augmentait, plus le nombre de fixations tendait également à augmenter. Dit autrement, moins les images étaient neutres (plus elles étaient soit positives, soit négatives), plus le nombre de fixations moyen tendait à augmenter. Ces résultats concordent avec ceux obtenus par (Carniglia et al., 2012), qui ont montré que les images émotionnellement chargées occasionnaient un plus grand nombre de fixations que les images neutres. Les conditions de visionnage n'étaient pas les mêmes dans nos deux études. Dans (Carniglia et al., 2012), les participants voyaient les images durant 12 secondes, et il leur était demandé de simplement regarder les images. Nos résultats sont intéressants en ceci qu'ils généralisent les résultats de (Carniglia et al., 2012) à des conditions différentes ; en particulier, des conditions très proches de celles dans lesquelles les scores de mémorabilité des images utilisées par l'ensemble des études s'étant intéressées à la prédiction de mémorabilité ont été collectés (Isola et al., 2011b).

Globalement, nos résultats expérimentaux confirment l'intérêt des caractéristiques des images liées à l'attention visuelle — et, par conséquent, de l'utilisation de modèles d'attention visuelle — pour la prédiction de la mémorabilité, en prenant en en compte ou non l'émotion véhiculée par les images.

Sur la performance des modèles d'attention visuelle

Nous avons montré, en utilisant des mesures de comparaison de cartes de saillance standard, que certains modèles étaient plus performants que d'autres sur l'ensemble des images utilisées. Il faut garder à l'esprit que la performance des modèles est probable-

ment relative au type de tâche réalisée par les participants dont les données oculométriques permettent d'obtenir la vérité terrain.

Nous avons trouvé que la performance des modèles tendait soit à augmenter avec le degré d'arousal des images (modèles *sim*, *resemb*, *rare*), soit au contraire à baisser (modèle *cov*), soit à rester stable (modèles *fast*, *itti* et *gbvs*). Ces résultats coïncident avec nos attentes, ainsi qu'avec nos résultats expérimentaux, qui corroborent l'hypothèse d'un lien entre l'attention visuelle et l'émotion véhiculée par une image. Puisque la performance de certains modèles varie en fonction de l'émotion véhiculée par les images, il peut être utile de choisir un modèle plutôt qu'un autre en fonction des images auxquelles on s'intéresse (par exemple, des images neutres, positives, ou négatives). Dans ce cas, nous pourrons nous aider des graphiques 11.4 et 11.5 pour éclairer notre choix.

Nous avons trouvé que la performance des modèles *cov* et *fast* tendait à décroître lorsque les scores de mémorabilité des images augmentaient. Il faut noter, cependant, que ces modèles sont globalement les plus performants de ceux évalués (en utilisant la méthode de mesure AUC). On aura donc tout intérêt à les utiliser, malgré ce fait, pour la prédiction de la mémorabilité des images. Mancas et Le Meur ont comparé plusieurs modèles d'attention visuelle pour l'extraction de caractéristiques des images liées à l'attention visuelle dans un objectif de prédiction de leur mémorabilité (Mancas and Le Meur, 2013). Parmi ces modèles, qui n'incluaient pas les modèles *cov* et *fast*, mais incluaient le modèle *gbvs*, ils ont trouvé que le modèle *rare* était le plus performant, au sens où la couverture de saillance des images calculée à partir des cartes de saillance générées par ce modèle était celle qui discriminait le mieux les images appartenant aux différentes classes de mémorabilité, « hautement mémorable », « moyennement mémorable » et « peu mémorable ». (Pour plus de détails, voir la section 4.1.1 du chapitre 4.) Or, dans notre expérience, le modèle *gbvs* montre de meilleures performances que le modèle *rare*. Cela suggère que les modèles d'attention visuelle les plus performants pour prédire la vérité terrain ne sont pas nécessairement les plus performants pour l'extraction de caractéristiques des images dans un objectif de prédiction de leur mémorabilité. Il serait intéressant de reproduire les résultats de Le Meur et Mancas en utilisant les modèles *cov* et *fast*.

D'autre part, la corrélation entre la performance des modèles *cov* et *fast* et les scores de mémorabilité des images était plus forte lorsqu'on s'intéressait aux scores de mémorabilité obtenus au second test de mémoire que lorsqu'on s'intéressait aux scores de mémorabilité obtenus au premier test de mémoire. Une explication possible est que, la tâche pour obtenir les scores de mémorabilité étant plus difficile pour le second test de mémoire que pour le premier, cela aura eu tendance à mieux répartir les scores de mémorabilité entre 0 (i.e. aucun participant n'a reconnu l'image) et 1 (i.e. tous les participants ont reconnu l'image), ce qui se sera traduit par une corrélation plus élevée pour le second test de mémoire que pour le premier.

Le fait que la variation de mémorabilité des images ou de l'arousal qu'elles véhiculent ne révèle de biais de performance que pour certains des modèles est intéressant. En effet, on pourrait imaginer qu'un certain nombre de facteurs que nous n'avons pas contrôlés, tels que le nombre d'objets dans l'image, le nombre de zones d'intérêt, le pourcentage de l'image couvert par ces zones, etc., sont corrélés à l'arousal véhiculé par les images ou à leur mémorabilité (malgré le fait que nous ayons sélectionné aléatoirement nos images dans l'IAPS). Or, le fait que seulement certains modèles montrent des différences de performance en fonction de l'arousal véhiculé par les images et de leur mémorabilité, suggère que ces différents facteurs que nous n'avons pas contrôlés n'ont pas eu d'influence indirecte significative sur nos résultats. Dans un tel cas de figure, ces facteurs auraient, en effet, influencé les performances de l'ensemble des modèles, et pas seulement de quelques uns.

Les modèles d'attention visuelle ont été également utilisés dans le cadre de la prédiction de la mémorabilité d'une manière différente de celle de (Mancas and Le Meur, 2013) (i.e. pour extraire des images des caractéristiques liées à l'attention visuelle qui ont une valeur prédictive de leur mémorabilité). Dans (Khosla et al., 2012b), les auteurs proposent un modèle probabiliste pour mesurer la mémorabilité de parties d'une image, qui repose, entre autres, sur l'utilisation de cartes de saillance générées par un modèle d'attention visuelle. Dans (Celikkale et al., 2015), les auteurs prédisent la mémorabilité des images à partir de caractéristiques de l'image, mais en se concentrant sur les régions saillantes de l'image, déterminées par un modèle d'attention visuelle. Il est probable que les études dans lesquelles sont utilisés des modèles d'attention visuelle pour la prédiction de la mémorabilité des images vont se multiplier.

11.5 Conclusion

Dans ce chapitre, nous avons montré que le nombre et la durée moyens des fixations occasionnées par le visionnage d'images sont liés aux émotions que celles-ci véhiculent et à leur mémorabilité. Ces résultats confirment l'intérêt de l'utilisation de caractéristiques des images liées à l'attention visuelle dans un cadre de prédiction de leur mémorabilité. De telles caractéristiques pouvant être computationnellement extraites des images à l'aide de modèles d'attention visuelle, nous nous sommes intéressés aux performances de plusieurs de ces modèles, en cherchant à déterminer si elles variaient conjointement avec la mémorabilité et la coloration émotionnelle des images. Nos résultats montrent que c'est effectivement le cas : certains modèles sont plus performants pour certains degrés de mémorabilité et d'émotion que d'autres. Ils ouvrent la porte à l'amélioration des modèles d'attention visuelle par l'intégration de ces facteurs. Dans une telle perspective, les progrès réalisés dans la prédiction de la mémorabilité des images et des émotions qu'elles suscitent, et ceux réalisés dans la modélisation de l'attention visuelle, pourraient se nourrir l'un l'autre. Un travail futur pourrait porter sur les caractéristiques

communes aux modèles d'attention visuelle, ou à leurs différences, qui expliquent l'influence des émotions véhiculées par les images et de leur mémorabilité sur la performance de certains modèles.

Le prochain et dernier chapitre traite d'un film interactif, dont le déroulement dépend des émotions ressenties par le spectateur. La mise en place de ce film a nécessité un enregistrement en temps réel des données oculométriques et des émotions des spectateurs, et une analyse conjointe de ces données.

12

Attention visuelle et réactions émotionnelles pour un film interactif

Le film interactif « émotionnel » est un film dont le contenu dépend en partie des réponses émotionnelles du spectateur. Il est né d'un partenariat initié par une réalisatrice, Marie-Laure Cazin[1], qui comprenait également Stéréolux, un espace culturel situé à Nantes, dédié aux musiques actuelles et aux arts numériques[2]. Marie-Laure Cazin désirait, pour son projet de « cinéma émotif »[3], un film dont le déroulement variât en fonction de l'émotion ressentie par les spectateurs. Dans le cadre de ce partenariat, notre intérêt était d'étudier la mémorisation d'un contenu multimédia en lien avec les émotions qu'il véhicule et le déploiement de l'attention visuelle des spectateurs.

Dans ce chapitre, nous décrivons le travail préliminaire qui a conduit à une version fonctionnelle du film interactif « émotionnel », et proposons le rôle que pourrait avoir cet outil pour élargir nos travaux sur la mémorabilité aux vidéos.

12.1 Introduction

L'art s'est souvent nourri des avancées technologiques. Les nouveaux outils offrent aux artistes de nouveaux moyens d'exprimer leur créativité. Aussi, les usages changent, et des formes originales d'interaction émergent entre les contenus artistiques et les indi-

[1] http ://www.marielaurecazin.net/
[2] http ://www.stereolux.org/
[3] https ://www.facebook.com/lecinemaemotif

vidus qui en font l'expérience. Le film interactif en est un exemple. Son déroulement dépend des réponses comportementales du spectateur. Une telle forme d'interaction propose une expérience renouvelée à l'utilisateur, et augmente potentiellement la qualité de son expérience et sa sensation d'immersion (Hu et al., 2005). En outre, elle offre aux créateurs l'opportunité d'augmenter leur contrôle sur l'expérience utilisateur avec la possibilité de définir les critères de l'interaction, c'est-à-dire de mieux faire respecter leur intention artistique.

Le « film interactif » est un terme générique qui regroupe différents concepts, pas nécessairement indépendants : le jeu interactif, un type de jeu vidéo essentiellement composé de scènes cinématiques et faisant un usage intensif des scripts (Hu et al., 2005) ; la vidéo interactive, une vidéo présentant des éléments cliquables ou actionnables (Schoeffmann et al., 2015) ; le cinéma interactif, dans lequel le public joue un rôle dans la manière dont le film se déroule (Hales, 2015) ; le film interactif, qui est une forme d'art interactive, dont l'objectif est de transformer les spectateurs en participants (actifs ou passifs) (Edmonds et al., 2004). Ce chapitre se concentre sur le dernier cas, et s'appuie sur la définition suivante : un film interactif est un film qui prend en compte les réactions des participants pour leur délivrer un contenu spécifique. Le film est découpé en différentes séquences, dont l'ajustement détermine le scénario cohérent particulier qui sera vu. L'articulation des séquences – dont résulte le scénario spécifique vu – est construite en interaction avec le spectateur, puisqu'elle dépend de ses réactions. À certains moments du film, dénommés embranchements, une séquence est sélectionnée parmi plusieurs possibles pour attribuer au film une structure spécifique.

Un film induit une cascade de réactions variées (comportementales, attentionnelles, émotionnelles, etc.) qui peuvent être mesurées à l'aide de différents outils (p. ex. caméra, oculomètre, EEG, etc.). Comme nous l'avons expliqué dans le chapitre 2, il y a aujourd'hui un fort intérêt de la communauté scientifique pour les réactions émotionnelles humaines, qui se traduit par un nombre grandissant d'études sur le sujet. En particulier, les émotions ont été envisagées comme une solution pour améliorer les interactions homme-technologie (Picard, 2010). Le film interactif émotionnel s'inscrit dans ce contexte : l'interaction film-spectateur repose sur les réactions émotionnelles du spectateur. Au-delà d'être une nouvelle forme de loisir numérique, ce film est une invitation à explorer une nouvelle manière de partager et de consommer des contenus multimédia.

12.2 Le fonctionnement du cinéma émotif

Dans cette section, nous décrivons le système utilisé dans le cinéma émotif, pour donner un aperçu de son fonctionnement. Dans la prochaine section, nous décrirons l'expérience réalisée en laboratoire pour parvenir à une version fonctionnelle du film interactif émotionnel. La figure 12.1 est une photographie du dispositif de cinéma émotif, prise lors d'une séance de présentation publique.

FIGURE 12.1 – Le projet de « cinéma émotif » dans lequel s'inscrit notre travail. Le spectateur est équipé avec un casque EEG Epoc. Un écran secondaire lui donne en temps réel un retour sur son propre état émotionnel.

12.2.1 Le dispositif EEG

Nous avons utilisé un casque d'EEG pour mesurer l'arousal des spectateurs pendant qu'ils regardaient le film. L'électroencéphalographie est une méthode d'exploration cérébrale que nous avons introduite dans cette thèse dans la section 2.3.2 du chapitre 2. Pour rappel, elle permet de mesurer en temps réel l'activité neurophysiologique. L'analyse des signaux enregistrés nous renseigne sur les processus cérébraux à l'œuvre (Saeid and Chambers, 2007).

L'intérêt de la communauté scientifique pour la reconnaissance automatique de l'émotion à partir des signaux EEG se manifeste, entre autre, dans le domaine des interfaces cerveau-machine, avec le développement de nouvelles formes d'interactions centrés sur l'humain (Liu et al., 2010c). Elle pourrait être utilisée dans des applications telles que l'interaction homme-machine, l'apprentissage assisté, la récupération d'informations, la santé humaine, la médecine préventive, ou encore les arts et loisirs (Picard and Picard, 1997). Des dispositifs d'EEG destinés au grand public, faciles à utiliser et relativement abordables, ont récemment été mis sur le marché. L'un d'entre eux est le casque Epoc d'Emotiv Systems[4], qui est utilisé dans le cinéma émotif. Le logiciel associé au casque EEG fournit directement des données émotionnelles sur plusieurs dimensions,

[4]https ://www.emotiv.com/epoc/

notamment sur la dimension d'arousal (appelée *Excitement*). La validité et la fiabilité du dispositif ont été vérifiées relativement aux données brutes qu'il fournit (Debener et al., 2012), et il a été montré que les données brutes pouvaient être utilisées pour la reconnaissance d'émotion (Pham and Tran, 2012). Cependant, à notre connaissance, aucune étude n'a évalué la validité des données émotionnelles que fournit le logiciel Emotiv associé au casque EEG.

12.2.2 Le film utilisé

Le film utilisé, « Mademoiselle Paradis », est un film à plusieurs scénarios qui a été écrit et réalisé par Marie-Laure Cazin. Il dure environ 20 minutes (selon le scénario déclenché), et comprend deux jonctions, avec quatre choix différents pour la première jonction et trois pour la seconde – pour un total de 12 scénarios différents (voir Figure 12.2). En dehors des séquences sélectionnables, le reste du film est identique pour tous les spectateurs. Quelque soit le déroulement du film, le scénario proposé à un spectateur est toujours cohérent. L'interprétation du film, en particulier concernant la responsabilité des différents personnages dans la maladie du personnage principal, Maria, qui constitue le cœur de l'intrigue, est susceptible de changer en fonction des séquences vues.

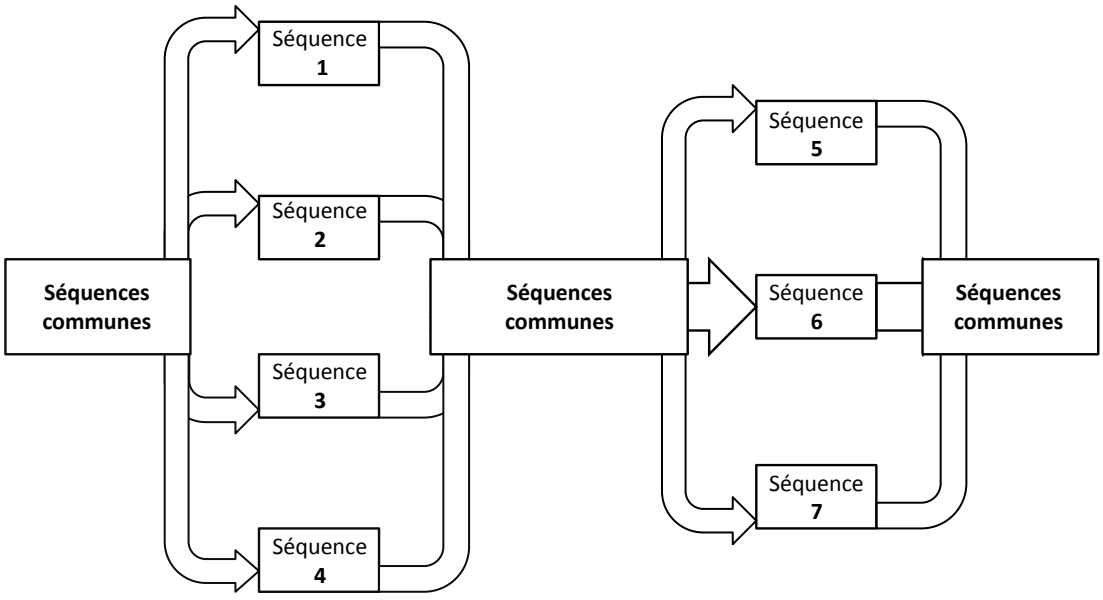


FIGURE 12.2 – Structure du film *Mademoiselle Paradis*, avec des séquences communes à tous les spectateurs et des séquences variables.

12.3 Une expérience préliminaire pour concevoir le système interactif

Afin de disposer d'un système entièrement fonctionnel pour les présentations publiques du cinéma émotif, nous avons conduit une expérience en laboratoire, que nous décrivons dans cette section.

12.3.1 Conception du système interactif

Plusieurs séquences de « Mademoiselle Paradis » sont, comme on l'a dit, susceptibles d'être choisies à l'une des deux jonctions. Une première nécessité était de définir un critère de sélection d'une séquence à une jonction. Pour ce premier prototype de film interactif émotionnel, nous avons décidé, à la suite des discussions avec la réalisatrice, que la séquence qui véhiculerait l'émotion la plus proche de l'état émotionnel du spectateur serait sélectionnée. L'idée était de suivre l'état émotionnel du spectateur pour prolonger son expérience émotionnelle. La mesure de l'état émotionnel correspondait aux données d'arousal fournies par le système EEG d'Emotiv. Le premier objectif de l'expérience était de déterminer l'arousal véhiculé par chacune des séquences sélectionnables à une jonction, de sorte que nous puissions déterminer, en comparant l'émotion qu'elles véhiculaient avec l'état émotionnel du spectateur, laquelle était la plus susceptible d'entrer en coïncidence avec celui-ci.

Deuxièmement, nous devions déterminer sur quelles parties du film l'état d'arousal des spectateurs serait enregistré pour effectuer le choix d'une séquence à un embranchement. Ces séquences d'enregistrement devaient satisfaire à plusieurs critères : (1) être temporellement proches des jonctions, pour que le choix de la séquence ne soit pas décorrélée de l'état émotionnel du spectateur au moment de l'embranchement, (2) ne pas être émotionnellement neutres, mais induire un arousal significatif, et (3) susciter des réactions suffisamment variées chez les différents spectateurs (i.e. qui permettent de discriminer les spectateurs pour adapter le scénario en fonction de leurs expériences propres). Concernant le premier point, nous avons, comme nous le verrons, sélectionné des parties du film proches des jonctions pour l'enregistrement de l'activité émotionnelle des spectateurs. En ce qui concerne le second point, nous avons utilisé les données émotionnelles fournies par le logiciel Emotiv associé au casque EEG pour caractériser les séquences du film en fonction de l'arousal qu'elles suscitaient. Pour satisfaire au troisième critère, nous avons utilisé un oculomètre afin de caractériser les séquences du film en fonction du comportement oculaire des spectateurs. Le logiciel de l'EEG que nous avons utilisé fournit des données émotionnelles à partir des enregistrements effectués par les électrodes du casque en utilisant un modèle boîte-noire : nous ne savons pas à quoi correspondent ces données. En particulier, il est possible que les variations des valeurs d'arousal fournies par le logiciel soient dues à quelque chose d'autre que

le contenu affiché (p. ex. le bruit produit par les mouvements du corps et enregistré par les électrodes). En utilisant un oculomètre, nous nous sommes assurés que les réactions étaient en relation avec le film affiché à l'écran. L'oculométrie ouvre, comme on l'a précédemment évoqué, une porte sur les processus cérébraux : sur les processus cognitifs (i.e. la prise de décision (Glöckner and Herbold, 2011), la mémoire (Mancas and Le Meur, 2013), le contrôle métacognitif (Roderer and Roebers, 2014)), et sur les processus émotionnels puisque la manière de regarder une image est liée à l'émotion qu'elle suscite (e.g. (Schupp et al., 2007)). Comme le comportement oculaire reflète les processus cérébraux sous-jacents, nous avons supposé que plus le comportement oculaire des spectateurs était varié, plus les traitements qu'ils effectuaient l'étaient également. Après avoir caractérisé les différentes parties du film en fonction de leur puissance discriminante, calculée à partir des données oculométriques des spectateurs, et de leur puissance émotionnelle, calculée à partir des données d'arousal fournies par le logiciel de l'EEG, nous avons déterminé quelles étaient les séquences-clés où enregistrer l'activité émotionnelle des spectateurs — c'est-à-dire, les séquences causant à la fois un arousal fort et des réactions variées.

12.3.2 Participants

Soixante participants (33 hommes et 27 femmes, de 19 à 61 ans, avec une moyenne de 24, 1 ans), rémunérés pour leur participation, ont pris part à l'étude. Tous les participants avaient une acuité visuelle normale ou corrigée. À leur arrivée, les participants passaient un test d'acuité visuel (les échelles Monoyer) et un test de vision des couleurs (test de Ishihara). Puis ils signaient un formulaire de consentement s'ils souhaitaient participer à l'expérience.

12.3.3 Matériel

Nous avons travaillé sur quatre version fixes du film *Mademoiselle Paradis* : la version A, composée des séquences 1 (pour le premier embranchement) et 5 (pour le second embranchement), B (séquences 2 et 6), C (séquences 3 et 7) et D (séquences 4 et 5). Toutes les séquences sélectionnables étaient vues par au moins un groupe de 15 spectateurs. Le film durait environ 20 minutes (en fonction de la version), avec le premier embranchement advenant à 7 minutes et 32 secondes et le second à environ 11 minutes. Nous ne nous sommes intéressés qu'aux 11 minutes qui précédaient le second embranchement. Les données d'EEG étaient acquises par les 14 électrodes du casque Emotiv, avec une fréquence de 128 Hz. Les mouvements oculaires étaient enregistrés simultanément à l'aide d'un oculomètre SMI RED. Le système comprend une caméra infrarouge et deux sources de lumière infrarouge, une de chaque côté de la caméra. L'oculomètre enregistre des points de regard (PG) à une fréquence de 50 PG/sec pour chaque œil.

12.3.4 Procédure

L'expérience avait lieu dans une salle d'expérience conçue et équipée pour l'évaluation de la qualité d'image (suivant les recommandations ITU-R BT.509). À leur arrivée, les participants étaient divisés entre quatre groupes, correspondant aux quatre versions (A, B, C et D) du film présentées. Ils s'installaient confortablement sur un siège, à une distance de $150cm$ (i.e. trois fois la hauteur de l'écran) d'un moniteur de 40 pouces (la résolution d'écran était de $1920x1080$), puis l'expérimentateur mettait en place le casque EEG. Une fois que chacune des électrodes renvoyait un signal optimal, une calibration en cinq points du système d'oculométrie était effectuée. L'expérimentateur lançait alors le film.

12.3.5 Résultats

Nous avons effectué une analyse des données collectées dans l'objectif de (1) sélectionner les parties du film qui, à la fois, induisaient un arousal significatif chez les spectateurs, et pour lesquelles la variabilité inter-observateur du comportement oculaire était importante, et (2) de caluler un « score d'arousal » pour chacune des sept séquences sélectionnables. En raison d'un enregistrement non optimal pour certains participants, l'analyse des données oculométriques et des données d'arousal dérivées des signaux EEG ont porté sur 59 et 52 participants, respectivement. De plus, nos analyses ont porté sur la description des *plans* du film ; ainsi avons-nous manuellement fractionné le film (ou plutôt : la partie du film précédent la seconde jonction) en plusieurs unités correspondant chacune à un plan, qui est la plus petite unité utilisée dans un film pour exprimer une émotion, une idée et/ou un mouvement particulier. Pour comparer nos variables (dont on verra qu'elles sont l'arousal moyen par plan et la dispersion moyenne par plan), nous les avons centrées et réduites.

Analyse des données oculométriques

À partir des données brutes fournies par le logiciel de l'oculomètre, nous avons extrait les points de regard des participants pour chaque image du film. La figure 12.3 donne un exemple des points de regard des différents spectateurs sur une image tirée d'une partie commune du film.

Nous avons ensuite calculé la dispersion D des points de regard pour chaque image x, tel que :

$$D_x = \frac{1}{n} \sum_{i=1}^{n} (d(p, c))_i$$

avec $d(p, c)_i$ la distance euclidienne entre le point de regard p_i et le barycentre c des n points de regard collectés pour l'image.

La dispersion moyenne des points de regard par plan y était calculée ainsi :

$$D_y = \frac{1}{n} \sum_{k=1}^{k} D_{x_k}$$

avec n le nombre d'images composant le plan.

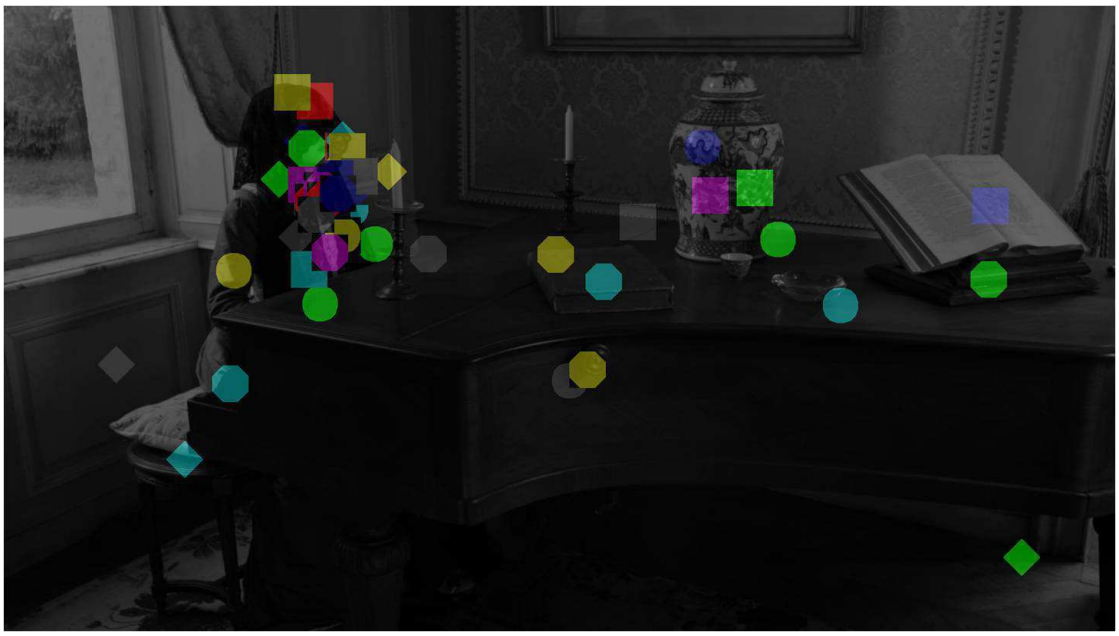


FIGURE 12.3 – Une image du film *Mademoiselle Paradis*. Chaque forme géométrique d'une couleur donnée correspond au point de regard d'un spectateur.

Analyse des données d'arousal

À partir des données d'arousal fournies par le système Emotiv, nous avons calculé pour chaque plan l'arousal moyen qu'il avait suscité chez l'ensemble des participant l'ayant vu. L'objectif était double : décrire chacune des sept séquences sélectionnables avec un simple score d'arousal, et déterminer quels plans induisaient un arousal suffisamment fort pour être intéressants pour l'enregistrement des réactions émotionnelles des spectateurs servant à sélectionner une séquence particulière à un embranchement.

Le score d'arousal A d'une séquence sélectionnable s, vue par n participants, pour chacun desquels m valeurs d'arousal ont été enregistrées, était calculé ainsi :

$$A_s = \frac{1}{n \times m} \sum_{i=1}^{n} \sum_{j=1}^{m} X_{i,j}$$

avec $X_{i,j}$ la j^{eme} valeur d'arousal fournie par le logiciel Emotiv pour le i^{eme} participant.

Les valeurs obtenues sont présentées dans la table 12.1.

Nous avons effectué le même calcul pour obtenir l'arousal moyen de chaque plan.

	Jonction 1				Jonction 2		
Numéro de la séquence	1	2	3	4	5	6	7
Score d'arousal	-0.69	-0.27	1.61	0.49	-1.31	1.05	0.02
Écart-type	0.27	0.15	0.24	0.21	0.38	0.26	0.26

TABLE 12.1 – Scores d'arousal moyens pour les séquence sélectionnables.

Parties du film discriminantes

Pour déterminer sur quelles parties du film l'activité émotionnelle des spectateurs serait enregistrée (i.e. les séquences dont nous attendons qu'elles suscitent à la fois une réaction émotionnelle forte et des réactions oculaires variées chez les spectateurs), nous avons placé sur le même graphique les scores d'arousal moyen et la dispersion moyenne des points de regard par plan (voir la figure 12.4). Nous avons choisi les plans associés à la fois à des scores d'arousal et à une dispersion des points de regard élevés pour enregistrer l'activité émotionnelle des spectateurs. Pour le premier embranchement (qui advient à la fin du plan 40), les plans sélectionnés sont les plans 35 à 37, pour une durée totale de 17 secondes. Pour le second embranchement (qui advient à la fin du plan 58), les plans sélectionnés sont les plans 47 à 49, pour une durée totale de 17 secondes. Il faut noter que les plans 47 à 49 sont décalés dans le temps entre les différentes version du film, ce qui doit être pris en compte dans le système opérationnel de film interactif.

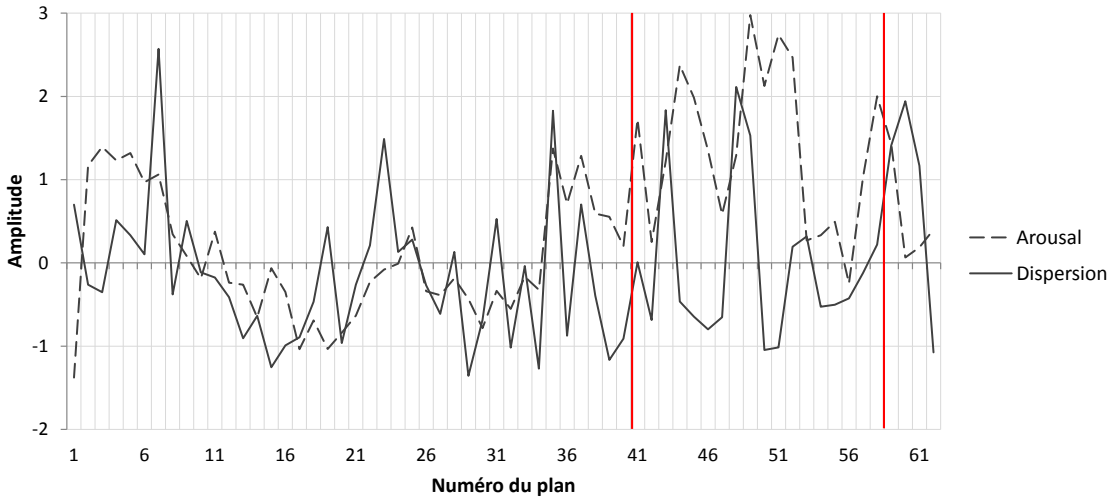


FIGURE 12.4 – Score d'arousal moyen et dispersion moyenne des points de regard par plan. Les données ont été centrées et réduites. Les lignes rouges verticales représentent les deux embranchements. Les parties du graphique correspondant aux séquences sélectionnables portent seulement sur les données des 15 participants ayant vue la version A du film.

Nous avons trouvé une corrélation linéaire positive entre les scores d'arousal et la dispersion moyenne par plan ($r = .30, p < .05$), ce qui signifie que l'arousal tel que mesuré par le casque Emotiv EEG n'est pas indépendant du contenu affiché (i.e. du film). Ce résultat suggère que la manière dont une personne regarde une scène est influencée par l'émotion qu'elle induit chez lui. Cette suggestion doit cependant être considérée avec prudence : en effet, comme nous l'avons précédemment évoqué, nous ne savons pas à quoi correspondent les données d'arousal que le logiciel Emotiv — une boîte noire — fournit ; par exemple, il est possible que les mouvements des yeux [5] enregistrés par les électrodes à la surface du scalp aient été pris en compte dans le calcul de l'arousal. D'autre part, nous n'avons pas contrôlé que les scènes qui suscitent le plus d'arousal ne soient pas également les scènes qui présentent les zones de saillance les plus dispersées. Par exemple, une interaction entre deux ou plusieurs personnes pourrait tendre à susciter une émotion plus forte qu'un personnage seul, tandis que le passage du regard d'un personnage à l'autre se traduirait par une augmentation de la dispersion des points de regard. Pour tester cette hypothèse, il serait intéressant de calculer des cartes de saillance à partir des données oculométriques, puis de les corréler avec l'arousal, pour se rendre compte dans quelle mesure l'arousal est lié à la dispersion des zones de saillance de l'image dans le film.

Il est également intéressant de noter que les scores d'arousal des plans sont positivement corrélés à leurs dispersions moyennes en matière de valeurs d'arousal : $r = .65, p < .001$. (La dispersion correspond ici aux écarts-types inter-sujets des valeurs d'arousal moyennes des spectateurs). Cela suggère que les séquences véhiculant les émotions les plus activatrices sont également celles qui véhiculent les réactions émotionnelles les plus variées.

Critère pour sélectionner une séquence à un embranchement

Nous avons défini le critère pour sélectionner une séquence comme suit : « la séquence dont le score d'arousal est le plus proche du score d'arousal du spectateur regardant le film (calculé en temps réel comme la moyenne des valeurs d'arousal pour les parties du film définies dans la section précédente) est sélectionnée ». Donc, la séquence sélectionnée SS à un embranchement est celle qui minimise la distance d, tel que :

$$SS_j = \arg\min_k d(A_{V_j}, A_{S_{j,k}})$$

avec A_V le score d'arousal du spectateur, A_S le score d'arousal de la séquence sélectionnable précédemment calculé, j l'indice de l'embranchement (1 ou 2) et k l'indice de la séquence sélectionnable.

[5] Les mouvements des yeux font partie des artéfacts bien connus dans les signaux EEG ; il convient généralement de les soustraire des analyses.

12.4 Discussion

Ce chapitre décrit une étude pilote dont l'objectif était de proposer une première version fonctionnelle d'un film qui change de contenu en fonction de l'émotion ressentie par le spectateur. Nous avons sélectionné des plans du film pendant la présentation desquels il nous a semblé pertinent d'enregistrer les réactions émotionnelles du spectateur, et proposé un critère pour adapter le contenu du film en fonction de celles-ci. Ce travail a été utilisé dans le « cinéma émotif » de Marie-Laure Cazin.

Des premières présentations publiques du cinéma émotif, il est apparu que certaines séquences sélectionnables tendaient à être plus souvent sélectionnées que d'autres. Cette inégalité n'était pas telle, cependant, que nous ayons décidé de la corriger pour les présentation publiques ultérieures, en pondérant la sélection des séquences sur la base de leurs probabilités d'occurrence. Dans la situation où plus d'un spectateur interagissaient avec le film, comme c'était le cas lors des présentations publiques du cinéma émotif, où deux spectateurs étaient équipés avec un casque EEG, la sélection d'une séquence à un embranchement dépendait d'un seul spectateur (qui changeait à l'embranchement suivant). Nous pouvons imaginer d'autres manières de procéder. Par exemple, dans le cadre d'un film vu en famille, on pourrait imaginer prendre en compte uniquement le spectateur qui ressent l'émotion la plus intense — ce qui pourrait augmenter l'immersion dans le film.

Une partie de notre étude avait pour objectif de caractériser les plans du films en fonction de l'arousal et des comportements oculaires qu'ils généraient chez un ensemble de spectateurs, en utilisant les données fournies par le dispositif EEG et l'oculomètre. La manière dont nous avons procédé a nécessité une expérience avec des participants ; pour éviter cette partie, d'autres méthodes pourraient être adoptées.

Plusieurs approches ont été proposées pour extraire computationnellement de l'information émotionnelle d'un film. L'approche proposée par (Baveye et al., 2015a), basée sur l'utilisation d'un réseau de neurones convolutifs, permet de calculer des scores d'arousal et de valence pour des portions de film d'une seconde (le modèle est décrit en détail dans (Baveye, 2015)). Une autre approche a été proposée par (Soleymani et al., 2009), qui repose sur une approche de classification bayésienne, et permet de classer les scènes d'un film dans trois catégories d'émotion différentes : « calme », « positivement excitante » et « négativement excitante ». Malandrakis *et al.* (Malandrakis et al., 2011) ont également proposé leur approche, basée sur un modèle de Markov caché, pour prédire l'émotion que les images d'une vidéo véhiculent. En utilisant une de ces méthodes, il est possible de qualifier automatiquement les séquences du film selon l'information émotionnelle qu'elles véhiculent (c'est-à-dire, sans avoir à passer par une expérience préliminaire nécessitant des participants). De même, les algorithmes de vision par ordinateur — tels que ceux utilisés dans le chapitre précédent (p. ex. GBVS (Harel et al., 2006)) — permettent de calculer la carte de saillance d'une image. Nous pourrions uti-

liser cette saillance visuelle prédite pour caractériser les parties du film en fonction de la probabilité qu'elles génèrent des comportements oculaires plus ou moins variés chez les spectateurs (en raison du plus ou moins grand nombre de régions d'intérêt).

Nous avons choisi une forme d'interaction avec le film basée sur l'arousal induit par le contenu : d'autres critères auraient pu être choisis. En lien avec le développement des interfaces neuronales directes, des méthodes ont été proposées pour inférer la valence de l'état émotionnel de signaux d'EEG (p. ex. (Schaaff, 2008, Horlings et al., 2008)). Si la mesure en temps réel de la valence est moins simple et probablement moins fiable que la mesure d'arousal, l'ajout de cette dimension de l'expérience utilisateur enrichirait l'interaction, et pourvoirait l'artiste d'un levier supplémentaire pour mieux transmettre son intention artistique. Nous pouvons ajouter qu'il est possible de faire correspondre à des valeurs sur les axes bipolaires d'arousal et de valence des états émotionnels discrets, tels que la tristesse, la relaxation, la joie, etc. (voir la figure 2.1 du chapitre 2). Par conséquent, ajouter l'information de valence permettrait au système d'interagir avec les spectateurs avec des mots qui auront certainement plus de sens pour eux que des degrés d'arousal et de valence.

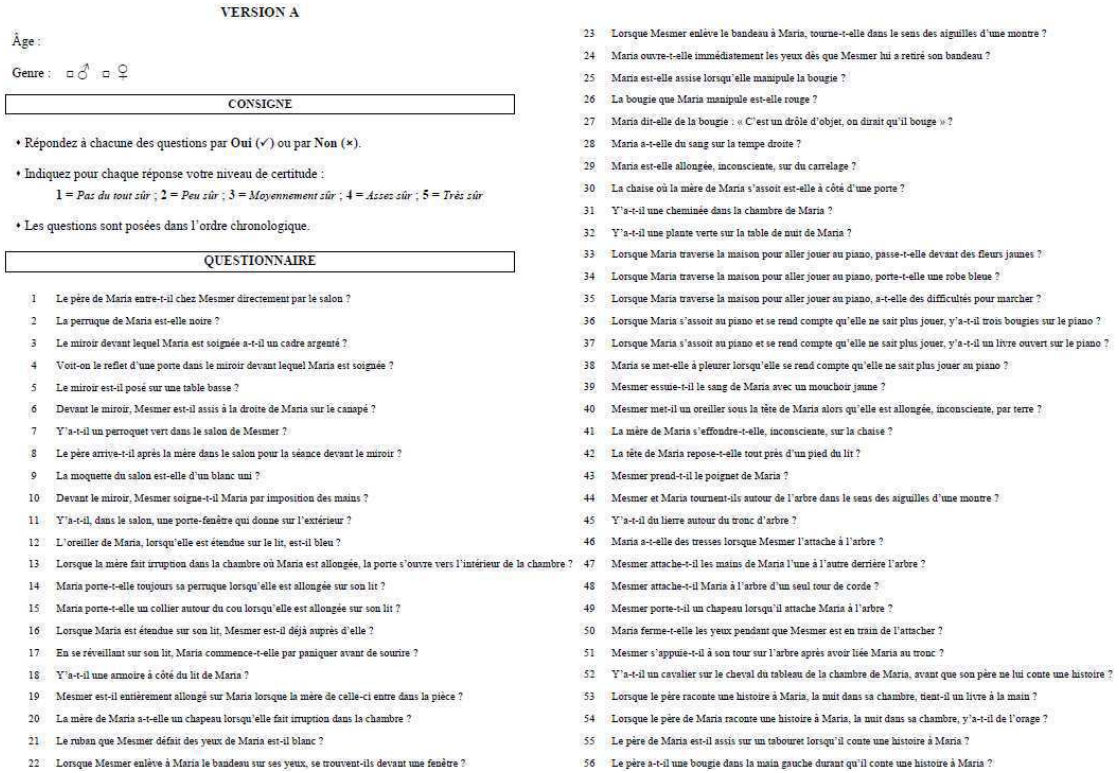
D'autres techniques de mesure pourraient également être employées pour enregistrer l'activité émotionnelle des spectateurs en temps réel (en particulier, les techniques présentées dans la section 2.3.2 du chapitre 2). La reconnaissance automatique des émotions est devenu un champ de recherche dynamique (Picard, 2010), et le développement de nouveaux outils pour recueillir et analyser des données émotionnelles avance rapidement. L'industrie du multimédia s'intéresse de près à de tels outils, pour au moins deux raisons, qui ont également participé à notre intérêt initial pour le film interactif émotionnel. La première raison est que ces outils pourraient permettre une personnalisation des contenus multimédias distribués. Généralement, les expériences multimédias sont optimisées pour un utilisateur moyen, comme nous l'avons expliqué dans la section 10.1.1 du chapitre 10. Les causes principales en sont la difficulté à prendre en compte les différences individuelles dans les modèles et à les mesurer de manière non intrusive. Le film interactif émotionnel est de ces systèmes qui pourraient permettre, par le moyen de la personnalisation, à la fois d'améliorer la qualité des expériences multimédias et d'optimiser le choix des contenus proposés à partir des retours individuels. La seconde raison est que ces outils pourraient permettre la création autonome de contenus numériques nouveaux ; par exemple, une machine pourrait piocher des contenu (images, sons, vidéos, etc.) dans des bases de données et les structurer à partir de retours utilisateur. C'est là, probablement, le prochain grand pas en avant pour l'art numérique — et le film interactif émotionnel est, en quelque sorte, un balbutiement.

12.5 Le principe du film interactif émotionnel pour l'étude de la mémorabilité des vidéos

Le film interactif émotionnel, ou plus généralement son principe, pourrait s'avérer intéressant pour élargir l'étude de la mémorabilité des images aux vidéos. Si une telle ouverture n'a pas, à notre connaissance, encore été proposée, nous pensons qu'elle ne tardera pas à susciter l'intérêt de notre communauté. Cet outil permet, en effet, d'étudier la mémorabilité en lien avec les réactions attentionnelles et émotionnelles individuelles. Il permet également de « jouer » avec l'expérience du spectateur (par exemple, en changeant le contexte émotionnel), pour déterminer dans quelles situations la mémorabilité du contenu multimédia — ou la qualité de l'expérience du spectateur — est maximisée.

Nous réfléchissons actuellement à une manière adaptée de mesurer la mémoire dans des vidéos. Nous avons profité de notre expérience en laboratoire et des présentations publiques du cinéma émotif pour faire passer une centaine de questionnaires de mémoire aux spectateurs, afin d'obtenir un première aperçu des comportements mnésiques en réaction au film. La figure 12.5 correspond au questionnaire la version A. Chaque questionnaire, qui correspondait à l'une ou l'autre des versions du film visionné, comprenait entre 56 et 60 questions à réponse binaire. Les questions étaient de trois types différents, correspondant aux trois éléments contextuels qui entrent dans la définition de la mémoire épisodique donnée par Tulving (donnée dans le chapitre 1 ; voir en particulier la section 1.2.3) : le « où », le « quand » et le « quoi ». La réponse à chacune des questions n'était apportée que par un seul plan du film, l'idée étant de croiser les résultats de mémoire avec les données émotionnelles et oculométriques des spectateurs enregistrées pendant qu'ils visionnaient la fraction du film correspondant à ce plan.

La description des résultats n'étant pas essentielle au regard de notre question de thèse, nous nous contentons de rapporter ici l'enseignement que nous en avons tiré. D'abord, il faut préciser que ce travail a été réalisé en début de thèse. Notre objectif était alors d'étudier la mémorisation d'un contenu multimédia en lien avec les émotions qu'il véhicule et le déploiement de l'attention visuelle des spectateurs. Les résultats non concluants (probablement parce que le test était trop difficile, et la formulation de certaines questions confusionnante) et les progrès dans notre étude de la littérature scientifique nous ont conduit à inscrire nos travaux dans le champ de recherche portant sur la mémorabilité des images. Ces non-résultats ont également contribué à notre intérêt pour les tâches de reconnaissance, en plus des raisons que nous avons évoquées dans la section 1.2.1 du chapitre 1. Notre vision des choses aujourd'hui est qu'une tâche de reconnaissance serait plus appropriée qu'un questionnaire de ce type pour l'étude de la mémorabilité des vidéos. Un telle tâche de reconnaissance porterait soit sur des images, ou des extraits, de vidéos. Nous pensons que les résultats obtenus à une tâche de reconnaissance de ce type nous permettrait d'étudier la prédiction computationnelle de la mémorabilité des vidéos d'une manière similaire à celle employée pour les images.



VERSION A

Âge :

Genre : ☐ ♂ ☐ ♀

CONSIGNE

- Répondez à chacune des questions par **Oui** (✓) ou par **Non** (×).
- Indiquez pour chaque réponse votre niveau de certitude :
  1 = *Pas du tout sûr* ; 2 = *Peu sûr* ; 3 = *Moyennement sûr* ; 4 = *Assez sûr* ; 5 = *Très sûr*
- Les questions sont posées dans l'ordre chronologique.

QUESTIONNAIRE

1 Le père de Maria entre-t-il chez Mesmer directement par le salon ?
2 La perruque de Maria est-elle noire ?
3 Le miroir devant lequel Maria est soignée a-t-il un cadre argenté ?
4 Voit-on le reflet d'une porte dans le miroir devant lequel Maria est soignée ?
5 Le miroir est-il posé sur une table basse ?
6 Devant le miroir, Mesmer est-il assis à la droite de Maria sur le canapé ?
7 Y'a-t-il un perroquet vert dans le salon de Mesmer ?
8 Le père arrive-t-il après la mère dans le salon pour la séance devant le miroir ?
9 La moquette du salon est-elle d'un blanc uni ?
10 Devant le miroir, Mesmer soigne-t-il Maria par imposition des mains ?
11 Y'a-t-il, dans le salon, une porte-fenêtre qui donne sur l'extérieur ?
12 L'oreiller de Maria, lorsqu'elle est étendue sur le lit, est-il bleu ?
13 Lorsque la mère fait irruption dans la chambre où Maria est allongée, la porte s'ouvre vers l'intérieur de la chambre ?
14 Maria porte-t-elle toujours sa perruque lorsqu'elle est allongée sur son lit ?
15 Maria porte-t-elle un collier autour du cou lorsqu'elle est allongée sur son lit ?
16 Lorsque Maria est étendue sur son lit, Mesmer est-il déjà auprès d'elle ?
17 En se réveillant sur son lit, Maria commence-t-elle par paniquer avant de sourire ?
18 Y'a-t-il une armoire à côté du lit de Maria ?
19 Mesmer est-il entièrement allongé sur Maria lorsque la mère de celle-ci entre dans la pièce ?
20 La mère de Maria a-t-elle un chapeau lorsqu'elle fait irruption dans la chambre ?
21 Le ruban que Mesmer défait des yeux de Maria est-il blanc ?
22 Lorsque Mesmer enlève à Maria le bandeau sur ses yeux, se trouvent-ils devant une fenêtre ?
23 Lorsque Mesmer enlève le bandeau à Maria, tourne-t-elle dans le sens des aiguilles d'une montre ?
24 Maria ouvre-t-elle immédiatement les yeux dès que Mesmer lui a retiré son bandeau ?
25 Maria est-elle assise lorsqu'elle manipule la bougie ?
26 La bougie que Maria manipule est-elle rouge ?
27 Maria dit-elle de la bougie : « C'est un drôle d'objet, on dirait qu'il bouge » ?
28 Maria a-t-elle du sang sur la tempe droite ?
29 Maria est-elle allongée, inconsciente, sur du carrelage ?
30 La chaise où la mère de Maria s'assoit est-elle à côté d'une porte ?
31 Y'a-t-il une cheminée dans la chambre de Maria ?
32 Y'a-t-il une plante verte sur la table de nuit de Maria ?
33 Lorsque Maria traverse la maison pour aller jouer au piano, passe-t-elle devant des fleurs jaunes ?
34 Lorsque Maria traverse la maison pour aller jouer au piano, porte-t-elle une robe bleue ?
35 Lorsque Maria traverse la maison pour aller jouer au piano, a-t-elle des difficultés pour marcher ?
36 Lorsque Maria s'assoit au piano et se rend compte qu'elle ne sait plus jouer, y'a-t-il trois bougies sur le piano ?
37 Lorsque Maria s'assoit au piano et se rend compte qu'elle ne sait plus jouer, y'a-t-il un livre ouvert sur le piano ?
38 Maria se met-elle à pleurer lorsqu'elle se rend compte qu'elle ne sait plus jouer au piano ?
39 Mesmer essuie-t-il le sang de Maria avec un mouchoir jaune ?
40 Mesmer met-il un oreiller sous la tête de Maria alors qu'elle est allongée, inconsciente, par terre ?
41 La mère de Maria s'effondre-t-elle, inconsciente, sur la chaise ?
42 La tête de Maria repose-t-elle tout près d'un pied du lit ?
43 Mesmer prend-t-il le poignet de Maria ?
44 Mesmer et Maria tournent-ils autour de l'arbre dans le sens des aiguilles d'une montre ?
45 Y'a-t-il du lierre autour du tronc d'arbre ?
46 Maria a-t-elle des tresses lorsque Mesmer l'attache à l'arbre ?
47 Mesmer attache-t-il les mains de Maria l'une à l'autre derrière l'arbre ?
48 Mesmer attache-t-il Maria à l'arbre d'un seul tour de corde ?
49 Mesmer porte-t-il un chapeau lorsqu'il attache Maria à l'arbre ?
50 Maria ferme-t-elle les yeux pendant que Mesmer est en train de l'attacher ?
51 Mesmer s'appuie-t-il à son tour sur l'arbre après avoir liée Maria au tronc ?
52 Y'a-t-il un cavalier sur le cheval du tableau de la chambre de Maria, avant que son père ne lui conte une histoire ?
53 Lorsque le père raconte une histoire à Maria, la nuit dans sa chambre, tient-il un livre à la main ?
54 Lorsque le père de Maria raconte une histoire à Maria, la nuit dans sa chambre, y'a-t-il de l'orage ?
55 Le père de Maria est-il assis sur un tabouret lorsqu'il conte une histoire à Maria ?
56 Le père a-t-il une bougie dans la main gauche durant qu'il conte une histoire à Maria ?

FIGURE 12.5 – Questionnaire de mémoire correspondant à la version A du film interactif émotionnel.

Conclusion

Dans cette partie, nous nous sommes intéressés à l'oculométrie et à la modélisation de l'attention visuelle pour l'étude de l'émotion véhiculée par les images et de leur mémorabilité. Nous avons également présenté un nouvel outil, le « film interactif émotionnel », qui repose sur un principe susceptible d'être intéressant pour élargir nos travaux aux vidéos.

Dans un premier temps, nous avons montré que l'attention visuelle était liée à l'émotion et à la mémorabilité des images, ce qui confirme l'intérêt de l'utilisation des modèles d'attention visuelle — qui permettent de calculer des caractéristiques d'une image liées à l'attention visuelle — pour la prédiction de la mémorabilité. Ensuite, nous avons mesuré la performance de sept modèles d'attention visuelle pour des images dont les scores d'émotion et de mémorabilité variaient. Nous avons trouvé que certains modèles présentent un bais : leur performance varie en fonction des scores d'émotion et/ou de mémorabilité des images. Ce résultat ouvre la porte à la prise en compte de ces dimensions pour améliorer la performance des modèles d'attention visuelle. Nous avons également détaillé les performances de chaque modèle en fonction du type d'émotion véhiculé par les images, ce qui permettra, le cas échéant, de choisir le modèle le plus performant pour une catégorie d'images en particulier.

Finalement, nous avons présenté une expérience qui nous a permis, à travers l'étude des données oculométriques et émotionnelles de spectateurs, de proposer une version fonctionnelle du film interactif émotionnel. Ces travaux ont été réalisés au début de cette thèse : ils ont participé à nous conduire à l'étude de la mémorabilité des images. Ils sont présentés à la fin de cette thèse : c'est que, nous voyons dans cet outil un potentiel pour élargir nos travaux sur la mémorabilité aux vidéos. L'ensemble des études présentées dans cette thèse — sur le développement d'un modèle de prédiction, l'étude des facteurs contextuels et individuels, ou les liens entre émotion, mémorabilité et attention visuelle — pourraient être adaptées à l'étude des vidéos à l'aide d'un tel outil. Dans le cinéma émotif, seul le scénario du film variait entre les différentes présentations ; nous pourrions imaginer manipuler d'autres facteurs, par exemple la qualité de l'image et du son, la luminosité, la taille d'affichage, etc., afin de déterminer les situations qui maximisent la mémorisation d'un contenu multimédia. Ces travaux laissent présager des perspectives intéressantes.

Dans le chapitre conclusif de cette thèse, nous revenons sur nos contributions à l'étude de la mémorabilité des images, et explorons quelques perspectives pour nos travaux futurs.

Conclusion générale et perspectives

Le questionnement qui a sous-tendu cette thèse était le suivant : comment améliorer la prédiction de la mémorabilité des images ? Pour répondre à cette question, nous avons commencé par renforcer l'assise théorique du champ de recherche dans lequel s'inscrit cette problématique. Dans cet objectif, nous avons opéré un rapprochement théorique entre la mémorabilité des images telle qu'elle est étudiée en vision par ordinateur et la mémoire humaine, étudiée en psychologie. Cela nous a permis de définir clairement ce qui était entendu par mémorabilité des images en vision par ordinateur. De plus, cet éclairage théorique a mis en lumière un certain nombre de voies, encore non exploitées, pour progresser dans sa prédiction. Les conclusions principales que nous tirons finalement de cette thèse — que la recherche sur la prédiction de la mémorabilité des images gagnerait à s'ouvrir à l'émotion véhiculée par les images, au contexte de leur présentation, à l'oubli en mémoire à long terme et à l'idiosyncrasie — doivent leur première esquisse à cette approche transversale que nous avons adoptée.

Les études réalisées durant cette thèse nous ont permis d'obtenir les meilleurs résultats de prédiction de la mémorabilité des images à ce jour, et de proposer de nouvelles pistes pour aller plus loin. Ainsi, la modélisation des liens entre plusieurs facteurs individuels et la mémorabilité des images que nous avons introduit est un premier travail en vue de personnaliser la prédiction de la mémorabilité des images. Nos travaux ont également indirectement confirmé l'intérêt de l'utilisation des modèles computationnels d'attention visuelle pour extraire des images des caractéristiques liées à l'attention visuelle et à leur mémorabilité. De plus, nous avons montré que certains modèles d'attention visuelle présentent un bais : leur performance varie en fonction des scores d'émotion et/ou de mémorabilité des images. Ce résultat ouvre la porte à la prise en compte de ces dimensions pour améliorer la performance des modèles d'attention visuelle.

Une des étapes importantes de ce travail de thèse a été la création d'une nouvelle base de données pour l'étude de la mémorabilité des images. Elle vient s'ajouter à l'unique base actuellement disponible, à notre connaissance. Elle permettra d'éprouver les modèles prédictifs sur de nouvelles images, ainsi que nous l'avons fait pour MemoNet — opération qui, en révélant un biais de notre modèle, dont la performance s'est avérée meilleure pour les images négatives et activatrices, nous a conduit à insister sur l'importance de la distribution dans l'espace émotionnel des images d'une base

de données destinée à l'étude de la mémorabilité. L'analyse des scores d'émotion et de mémorabilité que nous avons obtenus pour 150 images a confirmé deux points qui nous étaient apparus essentiels à la lumière de notre étude de la littérature. D'une part, les études sur la mémorabilité des images gagneraient à s'ouvrir aux émotions véhiculées par les images, en raison de l'étroite relations qu'elles entretiennent avec la mémorabilité des images. Cette ouverture serait facilitée par l'existence d'un champ de recherche parallèle au nôtre, qui porte sur l'extraction computationnel de l'information émotionnelle des images. D'autre part, si nous cherchons à prédire une mémorabilité pérenne, il est important considérer une mémorabilité calculée à partir de performances de mémoire mesurées après un délai de rétention mnésique suffisant, puisque la mémorabilité des images change au début de la mémoire à long terme. Cette analyse a également confirmé que l'annotation manuelle des images pour leur attacher des scores de mémorabilité était difficilement envisageable : MemoNet prédit mieux la mémorabilité des images que des humains.

L'étude réalisée au début de cette thèse qui a abouti à une version fonctionnelle du film interactif émotionnel, et qui, en lien avec notre étude de la littérature, a eu comme implication de nous conduire à étudier la mémorabilité des images, pourrait nous permettre de poursuivre nos travaux sur des vidéos. En nous dotant des moyens de manipuler les situations de mémorisation et d'enregistrer les réactions émotionnelles et attentionnelles des spectateurs, le principe sur lequel repose cet outil possède un potentiel intéressant pour une étude de la mémorabilité des vidéos qui engloberait les facteurs contextuels, émotionnels, individuels et attentionnels, pour lesquels nous avons dit notre intérêt. À notre connaissance, aucune tentative n'a encore été faite pour prédire la mémorabilité de vidéos, ou de parties de vidéos : c'est là une belle perspective pour des travaux futurs.

D'autres perspectives, pour la mémorabilité des images, nous semblent également intéressantes. Nous concluons cette thèse par leur présentation.

Prédiction de la qualité de la mémoire des images

Jusqu'à aujourd'hui, l'étude de la mémorabilité des images en informatique a, à notre connaissance, exclusivement porté sur la quantité d'informations récupérées en mémoire, mesurée à l'aide d'une tâche de reconnaissance. Comme nous l'avons expliqué dans la section 3.2.2 du chapitre 3, la qualité de la récupération mnésique peut également être considérée dans l'évaluation de la mémoire. Cette qualité renvoie à la puissance de la reviviscence du souvenir, qui s'appuie notamment sur les éléments contextuels associés à l'évènement mis en mémoire. La quantité des informations récupérées en mémoire et leur qualité ne concordent pas toujours. Les effets de l'émotion sur la mémoire en sont une bonne illustration : de nombreux effets de l'émotion sur la mémoire ne sont apparents que lorsque la qualité de la mémoire est considérée, et ces effets sont

également plus susceptibles d'advenir quand la vivacité de la mémoire est considérée (Kensinger and Schacter, 2008). Or, un des objectifs fondamentaux mis en avant par (Isola et al., 2014) pour la prédiction de la mémorabilité des images, est que, dans le futur, la compréhension informatique de la mémorabilité des images nous permette de crée des algorithmes qui acquière de la connaissance d'une manière similaire aux humains. Dans un tel objectif, l'élargissement de l'étude de la mémoire des images en informatique à la qualité des souvenirs d'images apparaît donc importante.

Dans le cadre d'une tâche de reconnaissance, ou pourra utiliser les paradigmes *Remember-Know* et *what-where-when* que nous avons présentés dans la section 1.2.3 du chapitre 1 pour évaluer la qualité de la mémoire. Ces techniques sont a priori compatibles avec le crowdsourcing.

Caractéristiques des images liées à la persistance de leur mémorabilité en MLT

Comme nous l'avons établi, la mémorabilité des images mesurée quelques minutes après leur encodage ne reflètent qu'en partie leur mémorabilité après une durée de rétention mnésique plus longue. L'oubli, qui ne frappe pas également tous les souvenirs d'images, modifie plus ou moins la mémorabilité des différentes images. Ainsi, l'arousal suscité par une image, probablement en raison de son influence sur les processus de consolidation mnésique, tend-il à la préserver de l'oubli. Il serait intéressant de chercher à lier l'oubli des différentes images à d'autres caractéristiques, pour comprendre ce qui fait qu'une image est oubliable en mémoire à long terme. Dans cet objectif, on pourra d'abord d'intéresser aux caractéristiques des images utilisées dans la prédiction de la mémorabilité des images. Le cadre de travail proposé initialement par (Isola et al., 2011b) pour la prédiction de la mémorabilité des images, présenté dans la section 4.3.1 du chapitre 4, est également adapté à un tel objectif. En pratique, en utilisant notre base de données, nous mettrions en sortie d'un SVM non plus les scores de mémorabilité des images, mais l'écart entre les scores de mémorabilité obtenus grâce à la première tâche de reconnaissance (qui mesurait la performance de mémoire quelques minutes après l'encodage des images) et les scores obtenus grâce à la seconde tâche de mémoire (qui mesurait la performance de mémoire un jour après l'encodage). De tels travaux compléteraient ceux portant sur la prédiction de la mémorabilité des images. D'autre part, ils permettraient, dans une certaine mesure, d'inférer des scores de mémorabilité des images de la base de (Isola et al., 2011b), calculés à partir de performances de mémoire mesurées quelques minutes après l'encodage des images, la mémorabilité de ces images après une période de rétention plus longue.

Intégration de l'idiosyncrasie dans les modèles prédictifs

Nous pensons qu'il est important de poursuivre les travaux sur l'intégration de l'idiosyncrasie dans les modèles prédictifs. La part de subjectivité dans la mémorabilité des images ne nous permettra pas de prédire celle-ci précisément si nous ne réalisons pas une telle intégration. Le modèle des liens entre plusieurs facteurs individuels et la probabilité de reconnaître une image que nous avons proposé dans le chapitre 10 se veut une ouverture vers la personnalisation de la prédiction de la mémorabilité des images. Cependant, une méthode plus efficace pourrait être de prendre en compte directement des informations individuelles, avec les images, en entrée d'un modèle à apprentissage profond. Il est probable qu'un tel objectif nécessite cependant un nombre conséquent de données. Pour cette raison, il est, à notre avis, important, que les auteurs de nouvelles bases de données rapportent le détails individuel des scores. Et pas seulement les auteurs de base de données destinées à l'étude de la mémorabilité des images. En effet, l'importance de l'intégration de l'idiosyncrasie dans les modèles prédictif dépasse le cadre de la prédiction de la mémorabilité d'image : il concerne aussi bien la prédiction des autres caractéristiques subjectives de l'image, telles que les émotions qu'elles véhiculent, leur intérêt, leur esthétique, leur qualité, etc. Plus généralement, la prise en compte des facteurs individuels par les modèles représente, à notre avis, un des défis essentiels que les chercheurs en traitement d'images devront relever.

Bibliographie

Abraham, W. C. and Robins, A. (2005). Memory retention–the synaptic stability versus plasticity dilemma. *Trends in neurosciences*, 28(2) :73–78.

Abrisqueta-Gomez, J., Bueno, O. F. A., Oliveira, M. G. M., and Bertolucci, P. H. F. (2002). Recognition memory for emotional pictures in Alzheimer's patients. *Acta Neurologica Scandinavica*, 105(1) :51–54.

Adam, S. (2003). Nouvelles techniques d'evaluation de la memoire : procedure de dissociation des processus et paradigme R/K. *Evaluation et prise en charge des troubles mnesiques*, pages 141–167.

Anderson, A. K. and Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, 411(6835) :305–309.

Ans, B. and Rousset, S. (2000). Neural networks with a self-refreshing memory : Knowledge transfer in sequential learning tasks without catastrophic forgetting. *Connection Science*, 12(1) :1–19.

Argembeau, A. D., Comblain, C., and Van der Linden, M. (2005). Affective valence and the self-reference effect : Influence of retrieval conditions. *British Journal of Psychology*, 96(4) :457–466.

Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory : A proposed system and its control processes. *The psychology of learning and motivation*, 2 :89–195.

Avero, P. and Calvo, M. G. (2006). Affective priming with pictures of emotional scenes : The role of perceptual similarity and category relatedness. *The Spanish journal of psychology*, 9(1) :10.

Backs, R. W., Silva, S. P. d., and Han, K. (2005). A Comparison of Younger and Older Adults' Self-Assessment Manikin Ratings of Affective Pictures. *Experimental Aging Research*, 31(4) :421–440.

Baddeley, A. (1986). *Oxford psychology series, No. 11. Working memory*. New York : Clarendon Press/Oxford University Press.

Baddeley, A. D. (1982). Implications of neuropsychological evidence for theories of normal memory. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 298(1089) :59–72.

Baddeley, A. D. (1997). *Human Memory : Theory and Practice*. Psychology Press.

Bainbridge, W. A., Isola, P., and Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology : General*, 142(4) :1323–1334.

Bargh, J. A., Chaiken, S., Govender, R., and Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, 62(6) :893–912.

Barrett, L. F. (2006). Are Emotions Natural Kinds ? *Perspectives on Psychological Science*, 1(1) :28–58.

Barrett, L. F. and Russell, J. A. (1999). The Structure of Current Affect Controversies and Emerging Consensus. *Current Directions in Psychological Science*, 8(1) :10–14.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4) :323–370.

Baveye, Y. (2015). *Automatic prediction of emotions induced by movies*. PhD thesis, Ecole Centrale de Lyon.

Baveye, Y., Dellandréa, E., Chamaret, C., and Chen, L. (2015a). Deep learning vs. kernel methods : Performance for emotion prediction in videos. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*, pages 77–83. IEEE.

Baveye, Y., Dellandrea, E., Chamaret, C., and Chen, L. (2015b). LIRIS-ACCEDE : A Video Database for Affective Content Analysis. *IEEE Transactions on Affective Computing*, 6(1) :43–55.

Bechara, A., Damasio, H., and Damasio, A. R. (2000). Emotion, Decision Making and the Orbitofrontal Cortex. *Cerebral Cortex*, 10(3) :295–307.

Bechara, A., Damasio, H., Damasio, A. R., and Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *The Journal of neuroscience*, 19(13) :5473–5481.

Bem, S. L. (1977). On the utility of alternative procedures for assessing psychological androgyny. *Journal of Consulting and Clinical Psychology*, 45(2) :196–205.

Bem, S. L. (1981). *Bem sex-role inventory*. Consulting Psychologists Press.

Berg, A. C., Berg, T. L., Daume, H., Dodge, J., Goyal, A., Han, X., Mensch, A., Mitchell, M., Sood, A., Stratos, K., and others (2012). Understanding and predicting importance in images. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3562–3569. IEEE.

Bernat, E., Patrick, C. J., Benning, S. D., and Tellegen, A. (2006). Effects of picture content and intensity on affective physiological response. *Psychophysiology*, 43(1) :93–103.

Berntson, G. G. and Cacioppo, J. T. (2000). From homeostasis to allodynamic regulation. *Handbook of psychophysiology*, 2 :459–481.

Blanchette, I. (2006). Snakes, spiders, guns, and syringes : How specific are evolutionary constraints on the detection of threatening stimuli ? *The Quarterly Journal of Experimental Psychology*, 59(8) :1484–1504.

Blaney, P. H. (1986). Affect and memory : A review. *Psychological Bulletin*, 99(2) :229–246.

Boiten, F. A. (1998). The effects of emotional behaviour on components of the respiratory cycle. *Biological psychology*, 49(1) :29–51.

Borkin, M., Vo, A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., and Pfister, H. (2013). What Makes a Visualization Memorable ? *IEEE Transactions on Visualization and Computer Graphics*, 19(12) :2306–2315.

Bowers, J. S. and Pleydell-Pearce, C. W. (2011). Swearing, euphemisms, and linguistic relativity. *PloS one*, 6(7) :e22341.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P. J. (2001). Emotion and motivation I : Defensive and appetitive reactions in picture processing. *Emotion*, 1(3) :276–298.

Bradley, M. M., Greenwald, M. K., Petry, M. C., and Lang, P. J. (1992). Remembering pictures : Pleasure and arousal in memory. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 18(2) :379–390.

Bradley, M. M. and Lang, P. J. (1994). Measuring emotion : The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1) :49–59.

Bradley, M. M. and Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In Coan, J. A. and Allen, J. J. B., editors, *Handbook of emotion elicitation and assessment*, Series in affective science., pages 29–46. Oxford University Press, New York, NY, US.

Bradley, M. M., Miccoli, L., Escrig, M. A., and Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4) :602–607.

Brady, T. F., Konkle, T., Alvarez, G. A., and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38) :14325–14329.

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

Braun, J. and Julesz, B. (1998). Withdrawing attention at little or no cost : Detection and discrimination tasks. *Perception & Psychophysics*, 60(1) :1–23.

Brosch, T., Scherer, K., Grandjean, D., and Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. *Swiss Medical Weekly*.

Buchanan, T. W. and Adolphs, R. (2002). The role of the human amygdala in emotional modulation of long-term declarative memory. In Moore, S. C. and Oaksford, M., editors, *Emotional cognition : From brain to behaviour*, Advances in Consciousness Research, vol. 44., pages 9–34. John Benjamins Publishing Company, Amsterdam, Netherlands.

Burke, A., Heuer, F., and Reisberg, D. (1992). Remembering emotional events. *Memory & cognition*, 20(3) :277–290.

Burt, D. B., Zembar, M. J., and Niederehe, G. (1995). Depression and memory impairment : A meta-analysis of the association, its pattern, and specificity. *Psychological Bulletin*, 117(2) :285–305.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., and Oliva, A. (2015a). Intrinsic and extrinsic effects on image memorability. *Vision Research*, 116, Part B :165–178.

Bylinskii, Z., Isola, P., Torralba, A., and Oliva, A. (2015b). Modeling Context Effects on Image Memorability.

Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., and Torralba, A. (2016). Mit saliency benchmark.

Cacioppo, J. T. and Gardner, W. L. (1999). Emotion. *Annual review of psychology*, 50(1) :191–214.

Cahill, L. and McGaugh, J. L. (1995). A Novel Demonstration of Enhanced Memory Associated with Emotional Arousal. *Consciousness and Cognition*, 4(4) :410–421.

Calvo, M. G., Nummenmaa, L., and Hyönä, J. (2007). Emotional and neutral scenes in competition : Orienting, efficiency, and identification. *The Quarterly Journal of Experimental Psychology*, 60(12) :1585–1593.

Carniglia, E., Caputi, M., Manfredi, V., Zambarbieri, D., and Pessa, E. (2012). The influence of emotional picture thematic content on exploratory eye movements. *Journal of Eye Movement Research*, 5(4) :1–9.

Carroll, J. M., M, S., Russell, J. A., and Barrett, L. F. (1999). On the psychometric principles of affect. *Review of General Psychology*, 3(1) :14–22.

Carver, C. S. and Scheier, M. F. (1990). Origins and functions of positive and negative affect : A control-process view. *Psychological Review*, 97(1) :19–35.

Cazin, M.-L. (https ://www.facebook.com/lecinemaemotif). Le cinema emotif.

Cazin, M.-L. (http ://www.marielaurecazin.net/). Site de Marie-Laure Cazin.

Celikkale, B., Erdem, A., and Erdem, E. (2013). Visual Attention-Driven Spatial Pooling for Image Memorability. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 976–983.

Celikkale, B., Erdem, A., and Erdem, E. (2015). Predicting memorability of images using attention-driven spatial pooling and image semantics. *Image and Vision Computing*, 42 :35–46.

Cernea, D., Kerren, A., and Ebert, A. (2011). Detecting insight and emotion in visualization applications with a commercial EEG headset. In *Proceedings of SIGRAD 2011. Evaluations of Graphics and Visualization—Efficiency ; Usefulness ; Accessibility ; Usability ; November 17-18 ; 2011 ; KTH ; Stockholm ; Sweden*, pages 53–60. Linköping University Electronic Press.

Chaby, L. and Narme, P. (2009). [processing facial identity and emotional expression in normal aging and neurodegenerative diseases]. *Psychologie & neuropsychiatrie du vieillissement*, 7(1) :31–42.

Chalupa, L. M. and Werner, J. S. (2004). *The visual neurosciences*. MIT press.

Charles, S. T., Mather, M., and Carstensen, L. L. (2003). Aging and emotional memory : The forgettable nature of negative images for older adults. *Journal of Experimental Psychology : General*, 132(2) :310–324.

Chen, M., Zhang, L., and Allebach, J. P. (2015). Learning deep features for image emotion classification. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4491–4495.

Choi, M. J., Lim, J. J., Torralba, A., and Willsky, A. S. (2010). Exploiting hierarchical context on a large database of object categories. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 129–136.

Christianson, S.-\., Loftus, E. F., Hoffman, H., and Loftus, G. R. (1991). Eye fixations and memory for emotional events. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 17(4) :693.

Christianson, S.-A. (1992). Emotional stress and eyewitness memory : A critical review. *Psychological Bulletin*, 112(2) :284–309.

Christianson, S.-A. (2014). *The Handbook of Emotion and Memory : Research and Theory*. Psychology Press.

Christianson, S.-A. and Fallman, L. (1990). The role of age on reactivity and memory for emotional pictures. *Scandinavian Journal of Psychology*, 31(4) :291–301.

Clayton, N. S. and Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, 395(6699) :272–274.

Clayton, N. S. and Dickinson, A. (1999). Scrub jays (Aphelocoma coerulescens) remember the relative time of caching as well as the location and content of their caches. *Journal of Comparative Psychology*, 113(4) :403–416.

Cohen, N. J. and Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia : dissociation of knowing how and knowing that. *Science*, 210(4466) :207–210.

Comblain, C., D'Argembeau, A., Linden, M. V. d., and Aldenhoff, L. (2004). The effect of ageing on the recollection of emotional and neutral pictures. *Memory*, 12(6) :673–684.

Coppin, G. and Sander, D. (2010). Théories et concepts contemporains en psychologie de l'émotion.

Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3) :201–215.

Cowan, N. (2012). *Working memory capacity*. Psychology press.

Crawley, S. and French, C. (2005). Field and observer viewpoint in remember-know memories of personal childhood events. *Memory*, 13(7) :673–681.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., and Lang, P. J. (2000). Brain potentials in affective picture processing : covariation with autonomic arousal and affective report. *Biological Psychology*, 52(2) :95–111.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1.

Dan-Glauser, E. S. and Scherer, K. R. (2011). The Geneva affective picture database (GAPED) : a new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, 43(2) :468–477.

Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2008). Image retrieval : Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, 40(2) :5.

De Graef, P. and Underwood, G. (2005). Semantic effects on object selection in real-world scene perception. *Cognitive processes in eye guidance*, pages 189–211.

Debener, S., Minow, F., Emkes, R., Gandras, K., and de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk ? *Psychophysiology*, 49(11) :1617–1621.

Deese, J. and Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of experimental psychology*, 54(3) :180.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet : A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE.

Dewhurst, S. A. and Parry, L. A. (2000). Emotionality, distinctiveness, and recollective experience. *European Journal of Cognitive Psychology*, 12(4) :541–551.

Dimberg, U. (1988). Facial electromyography and the experience of emotion. *Journal of Psychophysiology*.

Dolan, R. J. (2002). Emotion, Cognition, and Behavior. *Science*, 298(5596) :1191–1194.

Dolcos, F., LaBar, K. S., and Cabeza, R. (2004). Interaction between the Amygdala and the Medial Temporal Lobe Memory System Predicts Better Memory for Emotional Events. *Neuron*, 42(5) :855–863.

Dolcos, F., LaBar, K. S., and Cabeza, R. (2005). Remembering one year later : role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. *Proceedings of the national academy of sciences of the United States of America*, 102(7) :2626–2631.

Drace, S. (2013). Evidence for the role of affect in mood congruent recall of autobiographic memories. *Motivation and emotion*, 37(3) :623–628.

Duchowski, A. (2007). *Eye tracking methodology : Theory and practice*, volume 373. Springer Science & Business Media.

Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., and Dutoit, T. (2013). Performance of the Emotiv Epoc headset for P300-based applications. *BioMedical Engineering OnLine*, 12 :56.

Ebbinghaus, H. (1913). *Memory : A contribution to experimental psychology*. Number 3. University Microfilms.

Eccles, J. S., Jacobs, J. E., and Harold, R. D. (1990). Gender Role Stereotypes, Expectancy Effects, and Parents' Socialization of Gender Differences. *Journal of Social Issues*, 46(2) :183–201.

Edmonds, E., Turner, G., and Candy, L. (2004). Approaches to interactive art systems. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 113–117. ACM.

Eelen, J. D. H. D. H. P. (1998). Affective and Identity Priming with Episodically Associated Stimuli. *Cognition and Emotion*, 12(2) :145–169.

Eich, E., Macaulay, D., and Ryan, L. (1994). Mood dependent memory for events of the personal past. *Journal of Experimental Psychology : General*, 123(2) :201.

Eich, E. and Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 15(3) :443–455.

Ekanayake, H. (2010). *P300 and Emotiv EPOC : Does Emotiv EPOC capture real EEG (2010)*. Web publication.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4) :169–200.

Ekman, P. (1999). *Basic Emotions*. John Wiley & Sons, Ltd.

Ekman, P. and Friesen, W. V. (1977). Facial action coding system.

Ekman, P. and Friesen, W. V. (1978). Facial action coding system (facs) : Manual.

Emotiv, E. Headset.

Erdem, E. and Erdem, A. (2013). Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of vision*, 13(4) :11–11.

Ericsson and Qualcomm (2013). A focus on efficiency : A whitepaper from facebook. *Facebook*.

Eysenck, M. W. (1976). Arousal, learning, and memory. *Psychological Bulletin*, 83(3) :389–404.

Eysenck, M. W. (2014). Depth, elaboration, and distinctiveness. *Levels of Processing in Human Memory (PLE : Memory)*, 5 :89.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., and Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50(2) :229–238.

Fedorovskaya, E. A. and De Ridder, H. (2013). Subjective matters : from image quality to image psychology. In *IS&T/SPIE Electronic Imaging*, pages 86510O–86510O. International Society for Optics and Photonics.

Flykt, A. (2005). Visual search with biological threat stimuli : Accuracy, reaction times, and heart rate changes. *Emotion*, 5(3) :349–353.

Fox, E., Russo, R., Bowles, R., and Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety ? *Journal of Experimental Psychology : General*, 130(4) :681.

Frintrop, S., Rome, E., and Christensen, H. I. (2010). Computational Visual Attention Systems and Their Cognitive Foundations : A Survey. *ACM Trans. Appl. Percept.*, 7(1) :6 :1–6 :39.

Gardiner, J. M., Ramponi, C., and Richardson-Klavehn, A. (1998). Experiences of Remembering, Knowing, and Guessing. *Consciousness and Cognition*, 7(1) :1–26.

Gbèhounou, S., Lecellier, F., and Fernandez-Maloigne, C. (2012). Extraction of emotional impact in colour images. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2012, pages 314–319. Society for Imaging Science and Technology.

Gil, S. (2009). Comment etudier les emotions en laboratoire. *Revue electronique de psychologie sociale*, 4 :15–24.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587.

Glöckner, A. and Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions : Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, 24(1) :71–98.

Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Aistats*, volume 9, pages 249–256.

Google (2008). More than one trillion images indexed. *Google images*.

Graf, P. and Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 11(3) :501–518.

Green, D. and Swets, J. (1966). Signal detection theory and psychophysics. 1966. *New York*, 888 :889.

Grühn, D. and Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS) : Do ratings become more extreme with age ? *Behavior Research Methods*, 40(2) :512–521.

Grühn, D., Scheibe, S., and Baltes, P. B. (2007). Reduced negativity effect in older adults' memory for emotional pictures : The heterogeneity-homogeneity list paradigm. *Psychology and Aging*, 22(3) :644–649.

Gross, J. J. and Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9(1) :87–108.

Hadland, K. A., Rushworth, M. F. S., Gaffan, D., and Passingham, R. E. (2003). The effect of cingulate lesions on social behaviour and emotion. *Neuropsychologia*, 41(8) :919–931.

Hales, C. (2015). 3 Interactive Cinema in the Digital Age. *Interactive Digital Narrative : History, Theory and Practice*, page 36.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in cognitive sciences*, 5(9) :394–400.

Harel, J., Koch, C., and Perona, P. (2006). Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552.

Hart, R. P., Buchsbaum, D. G., Wade, J. B., Hamer, R. M., and Kwentus, J. A. (1987). Effect of sleep deprivation on first-year residents' response times, memory, and mood. *Academic Medicine*, 62(11) :940–2.

Hedden, T. and Gabrieli, J. D. E. (2004). Insights into the ageing mind : a view from cognitive neuroscience. *Nature Reviews Neuroscience*, 5(2) :87–96.

Henderson, J. M. (2007). Regarding scenes. *Current directions in psychological science*, 16(4) :219–222.

Hermans, D., Houwer, J. D., and Eelen, P. (1994). The affective priming effect : Automatic activation of evaluative information in memory. *Cognition and Emotion*, 8(6) :515–533.

Hermans, D., Houwer, J. D., and Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition and Emotion*, 15(2) :143–165.

Hermans, D., Vansteenwegen, D., and Eelen, P. (1999). Eye Movement Registration as a Continuous Index of Attention Deployment : Data from a Group of Spider Anxious Students. *Cognition and Emotion*, 13(4) :419–434.

Heuer, F. and Reisberg, D. (1990). Vivid memories of emotional events : The accuracy of remembered minutiae. *Memory & Cognition*, 18(5) :496–506.

Öhman, A., Flykt, A., and Esteves, F. (2001). Emotion drives attention : detecting the snake in the grass. *Journal of experimental psychology : general*, 130(3) :466.

Holland, S. M. and Smulders, T. V. (2011). Do humans use episodic memory to solve a What-Where-When memory task ? *Animal cognition*, 14(1) :95–102.

Homma, I. and Masaoka, Y. (2008). Breathing rhythms and emotions. *Experimental Physiology*, 93(9) :1011–1021.

Horlings, R., Datcu, D., and Rothkrantz, L. J. M. (2008). Emotion Recognition Using Brain Activity. In *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, CompSysTech '08, pages 6 :II.1–6 :1, New York, NY, USA. ACM.

Hu, J., Janse, M., and Kong, H.-j. (2005). User experience evaluation of a distributed interactive movie. In *HCI International*.

Hu, P., Stylos-Allan, M., and Walker, M. P. (2006). Sleep Facilitates Consolidation of Emotional Declarative Memory. *Psychological Science*, 17(10) :891–898.

Humphrey, K., Underwood, G., and 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.

Hunt, R. R. and Worthen, J. B. (2006). *Distinctiveness and memory*. Oxford University Press.

Isola, P., Parikh, D., Torralba, A., and Oliva, A. (2011a). Understanding the Intrinsic Memorability of Images. In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 24*, pages 2429–2437. Curran Associates, Inc.

Isola, P., Xiao, J., Parikh, D., Torralba, A., and Oliva, A. (2014). What Makes a Photograph Memorable ? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7) :1469–1482.

Isola, P., Xiao, J., Torralba, A., and Oliva, A. (2011b). What makes an image memorable ? In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 145–152.

Ito, T. and Cacioppo, J. (2005). Variations on a human universal : Individual differences in positivity offset and negativity bias. *Cognition and Emotion*, 19(1) :1–26.

Ito, T. A., Cacioppo, J. T., and Lang, P. J. (1998). Eliciting Affect Using the International Affective Picture System : Trajectories through Evaluative Space. *Personality and Social Psychology Bulletin*, 24(8) :855–879.

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10–12) :1489–1506.

Itti, L., Koch, C., and 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.

Izard, C. E. (1993). *The Differential Emotions Scale : DES IV-A ;[a Method of Measuring the Meaning of Subjective Experience of Discrete Emotions]*. University of Delaware.

Jacob, R. and Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory ? *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 16(3) :464–470.

Jennings, P. D., McGinnis, D., Lovejoy, S., and Stirling, J. (2000). Valence and Arousal Ratings for Velten Mood Induction Statements. *Motivation and Emotion*, 24(4) :285–297.

Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q.-T., Wang, J., Li, J., and Luo, J. (2011). Aesthetics and Emotions in Images. *IEEE Signal Processing Magazine*, 28(5) :94–115.

Jost, T., Ouerhani, N., Von Wartburg, R., Müri, R., and Hügli, H. (2005). Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding*, 100(1) :107–123.

Just, M. A. and Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, 8(4) :441–480.

Ke, Y., Tang, X., and Jing, F. (2006). The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 419–426. IEEE.

Kensinger, E. A. (2004). Remembering emotional experiences : The contribution of valence and arousal. *Reviews in the Neurosciences*, 15(4) :241–252.

Kensinger, E. A. (2010). Neuroimaging the formation and retrieval of emotional memories. *Brain mapping : New research*.

Kensinger, E. A., Brierley, B., Medford, N., Growdon, J. H., and Corkin, S. (2002). Effects of normal aging and Alzheimer's disease on emotional memory. *Emotion*, 2(2) :118–134.

Kensinger, E. A. and Corkin, S. (2003). Memory enhancement for emotional words : Are emotional words more vividly remembered than neutral words ? *Memory & Cognition*, 31(8) :1169–1180.

Kensinger, E. A. and Corkin, S. (2004). Two routes to emotional memory : distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9) :3310–3315.

Kensinger, E. A. and Schacter, D. L. (2006). Amygdala Activity Is Associated with the Successful Encoding of Item, But Not Source, Information for Positive and Negative Stimuli. *The Journal of Neuroscience*, 26(9) :2564–2570.

Kensinger, E. A. and Schacter, D. L. (2008). Memory and emotion. *Handbook of emotions*, 3 :601–617.

Khosla, A., Bainbridge, W. A., Torralba, A., and Oliva, A. (2013). Modifying the memorability of face photographs. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3200–3207.

Khosla, A., Raju, A. S., Torralba, A., and Oliva, A. (2015). Understanding and predicting image memorability at a large scale. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2390–2398.

Khosla, A., Xiao, J., Isola, P., Torralba, A., and Oliva, A. (2012a). Image Memorability and Visual Inception. In *SIGGRAPH Asia 2012 Technical Briefs*, SA '12, pages 35 :1–35 :4, New York, NY, USA. ACM.

Khosla, A., Xiao, J., Torralba, A., and Oliva, A. (2012b). Memorability of image regions. In *Advances in Neural Information Processing Systems*, pages 305–313.

Kim, E. Y., Kim, S.-j., Koo, H.-j., Jeong, K., and Kim, J.-i. (2005). Emotion-Based Textile Indexing Using Colors and Texture. In Wang, L. and Jin, Y., editors, *Fuzzy Systems and Knowledge Discovery*, number 3613 in Lecture Notes in Computer Science, pages 1077–1080. Springer Berlin Heidelberg. DOI : 10.1007/11539506_133.

Kim, J., Yoon, S., and Pavlovic, V. (2013). Relative Spatial Features for Image Memorability. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pages 761–764, New York, NY, USA. ACM.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36(14) :1.

Kleinginna Jr, P. R. and Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion*, 5(4) :345–379.

Kleinsmith, L. J. and Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology*, 65(2) :190–193.

Koch, C. and Ullman, S. (1987). Shifts in Selective Visual Attention : Towards the Underlying Neural Circuitry. In Vaina, L. M., editor, *Matters of Intelligence*, number 188 in Synthese Library, pages 115–141. Springer Netherlands. DOI : 10.1007/978-94-009-3833-5_5.

Konkle, T., Brady, T. F., Alvarez, G. A., and Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology : General*, 139(3) :558–578.

Koutstaal, W. and Schacter, D. L. (1997). Gist-Based False Recognition of Pictures in Older and Younger Adults. *Journal of Memory and Language*, 37(4) :555–583.

Kozlovskiy, S. A., Vartanov, A. V., Nikonova, E. Y., Pyasik, M. M., and Velichkovsky, B. M. (2012). The cingulate cortex and human memory processe. *Psychology in Russia : State of the art*, 5.

Krig, S. (2014). Ground Truth Data, Content, Metrics, and Analysis. In *Computer Vision Metrics*, pages 283–311. Springer.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Kučera, D. and Haviger, J. (2012). Using mood induction procedures in psychological research. *Procedia-Social and Behavioral Sciences*, 69 :31–40.

Kuppens, P., Tuerlinckx, F., Russell, J. A., and Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, 139(4) :917–940.

LaBar, K. S. and Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7(1) :54–64.

LaBar, K. S. and Phelps, E. A. (1998). Arousal-Mediated Memory Consolidation : Role of the Medial Temporal Lobe in Humans. *Psychological Science*, 9(6) :490–493.

Lahrache, S., El Qadi, A., and El Ouazzani, R. (2016). Bag-of-features for image memorability evaluation. *IET Computer Vision*.

Lang, P. J. (1995). The emotion probe : Studies of motivation and attention. *American Psychologist*, 50(5) :372–385.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). International affective picture system (IAPS) : Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, pages 39–58.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2008). International affective picture system (IAPS) : Affective ratings of pictures and instruction manual. *Technical report A-8*.

Lang, P. J., Bradley, M. M., Cuthbert, B. N., and others (1999). International affective picture system (IAPS) : Instruction manual and affective ratings. *The center for research in psychophysiology, University of Florida*.

Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., and Nangia, V. (1998). Emotional arousal and activation of the visual cortex : An fMRI analysis. *Psychophysiology*, 35(2) :199–210.

Larsen, R. J. and Diener, E. (1992). Promises and problems with the circumplex model of emotion. In *Emotion*, Review of personality and social psychology, No. 13., pages 25–59. Sage Publications, Inc, Thousand Oaks, CA, US.

Lazarus, R. S. (1991). Emotion and adaptation. 1991. *Cite en*, page 9.

Le Doux, J. (1996). *The Emotional Brain : The Mysterious Underpinnings of Emotional Life, ed*. Simon and Schuster.

Le Meur, O. and Baccino, T. (2013). Methods for comparing scanpaths and saliency maps : strengths and weaknesses. *Behavior research methods*, 45(1) :251–266.

Le Meur, O., Le Callet, P., Barba, D., and Thoreau, D. (2006). A coherent computational approach to model the bottom-up visual attention. *IEEE transactions on pattern analysis and machine intelligence*, 28 :802–817.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553) :436–444.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4) :541–551.

LeDoux, J. (1998). *The emotional brain : The mysterious underpinnings of emotional life*. Simon and Schuster.

Leonesio, R. J. and Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory ? *Journal of experimental psychology : Learning, Memory, and Cognition*, 16(3) :464.

Lewis, P. A. and Critchley, H. D. (2003). Mood-dependent memory. *Trends in Cognitive Sciences*, 7(10) :431–433.

Li, M., Chai, Q., Kaixiang, T., Wahab, A., and Abut, H. (2009). EEG emotion recognition system. In *In-vehicle corpus and signal processing for driver behavior*, pages 125–135. Springer.

Libkuman, T. M., Otani, H., Kern, R., Viger, S. G., and Novak, N. (2007). Multidimensional normative ratings for the International Affective Picture System. *Behavior Research Methods*, 39(2) :326–334.

Lieury, A. (2005). *Psychologie de la memoire : histoire, theories, experiences*. Dunod.

Lithari, C., Frantzidis, C. A., Papadelis, C., Vivas, A. B., Klados, M. A., Kourtidou-Papadeli, C., Pappas, C., Ioannides, A. A., and Bamidis, P. D. (2009). Are Females More Responsive to Emotional Stimuli ? A Neurophysiological Study Across Arousal and Valence Dimensions. *Brain Topography*, 23(1) :27–40.

Liu, L., Chen, R., Wolf, L., and Cohen-Or, D. (2010a). Optimizing photo composition. In *Computer Graphics Forum*, volume 29, pages 469–478. Wiley Online Library.

Liu, N., Dellandrea, E., Tellez, B., and Chen, L. (2010b). Reconnaissance de la semantique emotionnelle portee par les images basee sur la theorie de l'evidence. *month*.

Liu, Y., Sourina, O., and Nguyen, M. K. (2010c). Real-Time EEG-Based Human Emotion Recognition and Visualization. In *2010 International Conference on Cyberworlds (CW)*, pages 262–269.

Loftus, E. F., Loftus, G. R., and Messo, J. (1987). Some facts about" weapon focus.". *Law and Human Behavior*, 11(1) :55.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2) :91–110.

Lucassen, M. P., Gevers, T., and Gijsenij, A. (2010). Adding texture to color : quantitative analysis of color emotions. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2010, pages 5–10. Society for Imaging Science and Technology.

Luo, Y. and Tang, X. (2008). Photo and video quality evaluation : Focusing on the subject. In *European Conference on Computer Vision*, pages 386–399. Springer.

Machajdik, J. and Hanbury, A. (2010). Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, pages 83–92. ACM.

Madan, C. R., Caplan, J. B., Lau, C. S. M., and Fujiwara, E. (2012). Emotional arousal does not enhance association-memory. *Journal of Memory and Language*, 66(4) :695–716.

Malandrakis, N., Potamianos, A., Evangelopoulos, G., and Zlatintsi, A. (2011). A supervised approach to movie emotion tracking. In *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 2376–2379. IEEE.

Mancas, M. and Le Meur, O. (2013). Memorability of natural scenes : The role of attention. In *2013 20th IEEE International Conference on Image Processing (ICIP)*, pages 196–200.

Marchewka, A., Zurawski, L., Jednorog, K., and Grabowska, A. (2014). The Nencki Affective Picture System (NAPS) : Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, 46(2) :596–610.

Martin, M. A. and Metha, A. (1997). Recall of early childhood memories through musical mood induction. *The arts in Psychotherapy*, 24(5) :447–454.

Mather, M., Shafir, E., and Johnson, M. K. (2000). Misremembrance of Options Past : Source Monitoring and Choice. *Psychological Science*, 11(2) :132–138.

Mather, M. and Sutherland, M. R. (2011). Arousal-Biased Competition in Perception and Memory. *Perspectives on Psychological Science*, 6(2) :114–133.

Mauss, I. B. and Robinson, M. D. (2009). Measures of emotion : A review. *Cognition and Emotion*, 23(2) :209–237.

Mayer, J. D. and Gaschke, Y. N. (1988). The experience and meta-experience of mood. *Journal of Personality and Social Psychology*, 55(1) :102–111.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex : insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3) :419.

McGaugh, J. L. (1992). Affect, neuromodulatory systems, and memory storage. *The handbook of emotion and memory : Research and theory*, pages 245–268.

McGaugh, J. L. (2000). Memory–a Century of Consolidation. *Science*, 287(5451) :248–251.

McGaugh, J. L. and Roozendaal, B. (2002). Role of adrenal stress hormones in forming lasting memories in the brain. *Current opinion in neurobiology*, 12(2) :205–210.

Mehrabian, A. (1996). Pleasure-arousal-dominance : A general framework for describing and measuring individual differences in Temperament. *Current Psychology*, 14(4) :261–292.

Mikels, J. A., Larkin, G. R., Reuter-Lorenz, P. A., and Carstensen, L. L. (2005). Divergent trajectories in the aging mind : Changes in working memory for affective versus visual information with age. *Psychology and Aging*, 20(4) :542–553.

Miller, G. A. (1956). The magical number seven, plus or minus two : some limits on our capacity for processing information. *Psychological Review*, 63(2) :81–97.

Milner, B. (1966). Amnesia following operation on the temporal lobes. *Amnesia*, pages 109–133.

Mohammadi, G. and Vinciarelli, A. (2012). Automatic Personality Perception : Prediction of Trait Attribution Based on Prosodic Features. *IEEE Transactions on Affective Computing*, 3(3) :273–284.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of experimental psychology*, 64(5) :482.

Murray, L. A. (1999). Mood Congruence and Depressive Deficits in Memory : A Forced-recall Analysis. *Memory*, 7(2) :175–196.

Murray, N., Vanrell, M., Otazu, X., and Parraga, C. A. (2011). Saliency estimation using a non-parametric low-level vision model. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 433–440.

Nairne, J. S. (2006). Modeling Distinctiveness : Implications for General Memory Theory. In Hunt, R. R. and Worthen, J. B., editors, *Distinctiveness and Memory*, pages 26–46. Oxford University Press.

Nakayama, K. and Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, 29(11) :1631–1647.

Niu, Y., Todd, R. M., and Anderson, A. K. (2012). Affective salience can reverse the effects of stimulus-driven salience on eye movements in complex scenes.

Nummenmaa, L., Hyönä, J., and Calvo, M. G. (2006). Preferential selective attention to emotional pictures : An eye movement study. *Emotion*, 6 :257–268.

Nyklíček, I., Thayer, J. F., and P, J. (1997). Cardiorespiratory differentiation of musically-induced emotions. *Journal of Psychophysiology*, 11(4) :304–321.

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar ? The experience and process of recognizing feelings past. *Journal of Experimental Psychology : General*, 129(2) :242–261.

Oliva, A., Isola, P., Khosla, A., and Bainbridge, W. A. (2013). What makes a picture memorable ? *SPIE Newsroom*.

Oliva, A. and Torralba, A. (2001). Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, 42(3) :145–175.

Ortony, A., Turner, T. J., and Antos, S. J. (1983). A puzzle about affect and recognition memory. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 9(4) :725–729.

Ou, L.-C., Luo, M. R., Woodcock, A., and Wright, A. (2004). A study of colour emotion and colour preference. Part I : Colour emotions for single colours. *Color Research & Application*, 29(3) :232–240.

Partala, T. and Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1–2) :185–198.

Payne, J. D., Ellenbogen, J. M., Walker, M. P., and Stickgold, R. (2008a). The Role of Sleep in Memory Consolidation. In *Learning and Memory : A Comprehensive Reference*, pages 663–685. Academic Press, Oxford.

Payne, J. D. and Kensinger, E. A. (2011). Sleep leads to changes in the emotional memory trace : evidence from fMRI. *Journal of Cognitive Neuroscience*, 23(6) :1285–1297.

Payne, J. D., Stickgold, R., Swanberg, K., and Kensinger, E. A. (2008b). Sleep preferentially enhances memory for emotional components of scenes. *Psychological Science*, 19(8) :781–788.

Pettinelli, M. (2008). *The psychology of emotions, feelings and thoughts*. Lulu. com.

Pham, T. D. and Tran, D. (2012). Emotion Recognition Using the Emotiv EPOC Device. In Huang, T., Zeng, Z., Li, C., and Leung, C. S., editors, *Neural Information Processing*, number 7667 in Lecture Notes in Computer Science, pages 394–399. Springer Berlin Heidelberg. DOI : 10.1007/978-3-642-34500-5_47.

Picard, R. W. (2010). Affective Computing : From Laughter to IEEE. *IEEE Transactions on Affective Computing*, 1(1) :11–17.

Picard, R. W. and Picard, R. (1997). *Affective computing*, volume 252. MIT press Cambridge.

Plancher, G., Gyselinck, V., Nicolas, S., and Piolino, P. (2010). Age effect on components of episodic memory and feature binding : A virtual reality study. *Neuropsychology*, 24(3) :379–390.

Plancher, G., Nicolas, S., and Piolino, P. (2008). Apport de la realite virtuelle en neuropsychologie de la memoire : etude dans le vieillissement. *Psychologie & Neuro-Psychiatrie du vieillissement*, 6(1) :7–22.

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect : An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(03) :715–734.

Pourtois, G., Schettino, A., and Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention : What is magic and what is not. *Biological Psychology*, 92(3) :492–512.

Quirk, S. W. and Strauss, M. E. (2001). Visual exploration of emotion eliciting images by patients with schizophrenia. *The Journal of nervous and mental disease*, 189(11) :757–765.

Rawson, K. A. and Van Overschelde, J. P. (2008). How does knowledge promote memory ? The distinctiveness theory of skilled memory. *Journal of Memory and Language*, 58(3) :646–668.

Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task : An attentional blink ? *Journal of Experimental Psychology : Human Perception and Performance*, 18(3) :849–860.

Revelle, W. and Loftus, D. A. (1992). The implications of arousal effects for the study of affect and memory. *The handbook of emotion and memory : Research and theory*, pages 113–149.

Revlin, R. (2012). *Cognition : Theory and Practice*. Palgrave Macmillan.

Ribeiro, R. L., Pompeia, S., and Bueno, O. F. A. (2005). Comparison of Brazilian and American norms for the International Affective Picture System (IAPS). *Revista Brasileira De Psiquiatria (Sao Paulo, Brazil : 1999)*, 27(3) :208–215.

Richardson-Klavehn, A. and Bjork, R. A. (1988). Measures of memory. *Annual review of psychology*, 39(1) :475–543.

Riche, N., Mancas, M., Duvinage, M., Mibulumukini, M., Gosselin, B., and Dutoit, T. (2013). RARE2012 : A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing : Image Communication*, 28(6) :642–658.

Riemann, B. C. and McNally, R. J. (1995). Cognitive processing of personally relevant information. *Cognition and Emotion*, 9(4) :325–340.

Roderer, T. and Roebers, C. M. (2014). Can you see me thinking (about my answers) ? Using eye-tracking to illuminate developmental differences in monitoring and control skills and their relation to performance. *Metacognition and Learning*, 9(1) :1–23.

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2007). LabelMe : A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision*, 77(1-3) :157–173.

Russell, J. A. and Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion : Dissecting the elephant. *Journal of Personality and Social Psychology*, 76(5) :805–819.

Saeid, S. and Chambers, J. A. (2007). EEG signal processing. *Chichester : John Willey & Sons. Ltd*.

Salvucci, D. D. and Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM.

Scapinello, K. F. and Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science*, 21(6) :329–330.

Schaaff, K. (2008). EEG-based emotion recognition. *Universitat Karlsruhe (TH)*.

Schacter, D. L. (1985). Priming of old and new knowledge in amnesic patients and normal subjects. *Annals of the New York Academy of Sciences*, 444 :41–53.

Schaefer, A., Nils, F., Sanchez, X., and Philippot, P. (2005). A multi-criteria assessment of emotional films. *Manuscript submitted for publication*.

Scherer, K. R. (2005). What are emotions ? and how can they be measured ? *Social science information*, 44(4) :695–729.

Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 11(3) :565.

Schoeffmann, K., Hudelist, M. A., and Huber, J. (2015). Video Interaction Tools : A Survey of Recent Work. *ACM Comput. Surv.*, 48(1) :14 :1–14 :34.

Schupp, H. T., Stockburger, J., Codispoti, M., Junghöfer, M., Weike, A. I., and Hamm, A. O. (2007). Selective visual attention to emotion. *The Journal of Neuroscience*, 27(5) :1082–1089.

Schwabe, L. and Wolf, O. T. (2010). Emotional modulation of the attentional blink : Is there an effect of stress ? *Emotion*, 10(2) :283–288.

Seo, H. J. and Milanfar, P. (2009). Static and space-time visual saliency detection by self-resemblance. *Journal of Vision*, 9(12) :15–15.

Shallice, T. and Warrington, E. K. (1970). Independent functioning of verbal memory stores : A neuropsychological study. *Quarterly Journal of Experimental Psychology*, 22(2) :261–273.

Shapiro, D. and Herbert, P. (1967). Arousal correlates of task role and group setting. *Journal of Personality and Social Psychology*, 5(1) :103–107.

Sharot, T., Delgado, M. R., and Phelps, E. A. (2004). How emotion enhances the feeling of remembering. *Nature Neuroscience*, 7(12) :1376–1380.

Sharot, T. and Phelps, E. A. (2004). How arousal modulates memory : Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, 4(3) :294–306.

Sharot, T. and Yonelinas, A. P. (2008). Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition*, 106(1) :538–547.

Shechtman, E. and Irani, M. (2007). Matching Local Self-Similarities across Images and Videos. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Sheikh, H. R., Bovik, A. C., and Cormack, L. (2005). No-reference quality assessment using natural scene statistics : Jpeg2000. *IEEE Transactions on Image Processing*, 14(11) :1918–1927.

Shi, Y., Ruiz, N., Taib, R., Choi, E., and Chen, F. (2007). Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*, pages 2651–2656. ACM.

Silva, H. P. D., Fairclough, S., Holzinger, A., Jacob, R., and Tan, D. (2015). Introduction to the special issue on physiological computing for human-computer interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(6) :29.

Simola, J., Le Fevre, K., Torniainen, J., and Baccino, T. (2015). Affective processing in natural scene viewing : Valence and arousal interactions in eye-fixation-related potentials. *NeuroImage*, 106 :21–33.

Slowiaczek, M. L. and Clifton, C. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning and Verbal Behavior*, 19(5) :573–582.

Smith, C. A. and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48(4) :813–838.

Soares, A. P., Pinheiro, A. P., Costa, A., Frade, C. S., Comesaña, M., and Pureza, R. (2014). Adaptation of the International Affective Picture System (IAPS) for European Portuguese. *Behavior Research Methods*, 47(4) :1159–1177.

Soleymani, M., Kierkels, J. J., Chanel, G., and Pun, T. (2009). A bayesian framework for video affective representation. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–7. IEEE.

Spence, J. T., Helmreich, R., and Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology*, 32(1) :29–39.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs : General and Applied*, 74(11) :1–29.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, 25(2) :207–222.

Stebbins, C. M. and Coolidge, M. H. (1909). *Golden Treasury Readers : Primer, [First-fifth Reader]*. American Book Company.

Steinmetz, K. R. M., Schmidt, K., Zucker, H. R., and Kensinger, E. A. (2012). The effect of emotional arousal and retention delay on subsequent-memory effects. *Cognitive Neuroscience*, 3(3-4) :150–159.

Stevenson, I., Duncan, C. H., and Ripley, H. S. (1950). Variations in the electrocardiogram changes in emotional state. *Geriatrics*, 6(3) :164–178.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437(7063) :1272–1278.

Stytsenko, K., Jablonskis, E., and Prahm, C. (2011). Evaluation of consumer EEG device Emotiv EPOC. In *MEi : CogSci Conference 2011, Ljubljana*.

Suja, P., Thomas, S. M., Tripathi, S., and Madan, V. K. (2016). Emotion Recognition from Images Under Varying Illumination Conditions. In *Soft Computing Applications*, pages 913–921. Springer.

Sweller, J. (1988). Cognitive Load During Problem Solving : Effects on Learning. *Cognitive Science*, 12(2) :257–285.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9.

Szummer, M. and Picard, R. W. (1998). Indoor-outdoor image classification. In *Content-Based Access of Image and Video Database, 1998. Proceedings., 1998 IEEE International Workshop on*, pages 42–51. IEEE.

Tatler, B. W., Hayhoe, M. M., Land, M. F., and Ballard, D. H. (2011). Eye guidance in natural vision : Reinterpreting salience. *Journal of Vision*, 11(5) :5–5.

Tavakoli, H. R., Rahtu, E., and Heikkilä, J. (2011). Fast and Efficient Saliency Detection Using Sparse Sampling and Kernel Density Estimation. In Heyden, A. and Kahl, F., editors, *Image Analysis*, number 6688 in Lecture Notes in Computer Science, pages 666–675. Springer Berlin Heidelberg. DOI : 10.1007/978-3-642-21227-7_-62.

Thompson, E. R. (2007). Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, 38(2) :227–242.

Todd, R. M., Talmi, D., Schmitz, T. W., Susskind, J., and Anderson, A. K. (2012). Psychophysical and Neural Evidence for Emotion-Enhanced Perceptual Vividness. *The Journal of Neuroscience*, 32(33) :11201–11212.

Toet, A. (2011). Computational versus psychophysical bottom-up image saliency : A comparative evaluation study. *IEEE transactions on pattern analysis and machine intelligence*, 33(11) :2131–2146.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1) :97–136.

Trezise, K. and Reeve, R. A. (2014). Cognition-emotion interactions : patterns of change and implications for math problem solving. *Frontiers in Psychology*, 5.

Tsai, J. L., Knutson, B., and Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology*, 90(2) :288–307.

Tulving, E. (1972). Episodic and semantic memory 1. *Organization of Memory. London : Academic*, 381(4).

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, 26(1) :1.

Ucros, C. G. (1989). Mood state-dependent memory : A meta-analysis. *Cognition and Emotion*, 3(2) :139–169.

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, 71(5) :673–679.

Union, I. T. P.800 : Methods for subjective determination of transmission quality.

Upadhyaya, N. and Dixit, M. (2016). A review : Relating low level features to high level semantics in cbir. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 9(3) :433–444.

Valenti, R., Sebe, N., Gevers, T., et al. (2007). Facial expression recognition : A fully integrated approach. In *Proceedings of the 14th international conference of image analysis and processing-workshops*, pages 125–130. IEEE Computer Society.

Valenza, G., Lanata, A., and Scilingo, E. P. (2012). The Role of Nonlinear Dynamics in Affective Valence and Arousal Recognition. *IEEE Transactions on Affective Computing*, 3(2) :237–249.

Vassilieva, N. S. (2009). Content-based image retrieval methods. *Programming and Computer Software*, 35(3) :158–180.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., and Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition and Emotion*, 22(4) :720–752.

Walker, M. P. (2009). The Role of Sleep in Cognition and Emotion. *Annals of the New York Academy of Sciences*, 1156(1) :168–197.

Wang, W., Sun, J., Li, J., Wu, Q., and Liu, J. (2015). Investigation on the Influence of Visual Attention on Image Memorability. In Zhang, Y.-J., editor, *Image and Graphics*, number 9219 in Lecture Notes in Computer Science, pages 573–582. Springer International Publishing. DOI : 10.1007/978-3-319-21969-1_52.

Wang, W.-N. and Yu, Y.-L. (2005). Image emotional semantic query based on color semantic description. In *Proceedings of 2005 International Conference on Machine Learning and Cybernetics, 2005*, volume 7, pages 4571–4576 Vol. 7.

Watkins, P. C., Mathews, A., Williamson, D. A., and Fuller, R. D. (1992). Mood-congruent memory in depression : Emotional priming or elaboration ? *Journal of Abnormal Psychology*, 101(3) :581–586.

Watkins, P. C., Vache, K., Verney, S. P., and Mathews, A. (1996). Unconscious mood-congruent memory bias in depression. *Journal of Abnormal Psychology*, 105(1) :34.

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect : the panas scales. *Journal of personality and social psychology*, 54(6) :1063.

Wei, K., He, B., Zhang, T., and He, W. (2008). Image emotional classification based on color semantic description. In *International Conference on Advanced Data Mining and Applications*, pages 485–491. Springer.

Weymar, M., Löw, A., and Hamm, A. O. (2011). Emotional memories are resilient to time : Evidence from the parietal ERP old/new effect. *Human Brain Mapping*, 32(4) :632–640.

Wheeler, M. A., Stuss, D. T., and Tulving, E. (1997). Toward a theory of episodic memory : The frontal lobes and autonoetic consciousness. *Psychological Bulletin*, 121(3) :331–354.

White, R. (2002). Memory for events after twenty years. *Applied Cognitive Psychology*, 16(5) :603–612.

Witvliet, C. V. and Vrana, S. R. (1996). The emotional impact of instrumental music on affect ratings, facial EMG, autonomic measures, and the startle reflex : Effects of valence and arousal. In *Psychophysiology*, volume 33, pages S91–S91. Soc Psychophysiol Res 1010 Vermont Ave NW Suite 1100, Washington, DC 20005.

Wrase, J., Klein, S., Gruesser, S. M., Hermann, D., Flor, H., Mann, K., Braus, D. F., and Heinz, A. (2003). Gender differences in the processing of standardized emotional visual stimuli in humans : a functional magnetic resonance imaging study. *Neuroscience Letters*, 348(1) :41–45.

Wulfert, E., Roland, B. D., Hartley, J., Wang, N., and Franco, C. (2005). Heart rate arousal and excitement in gambling : Winners versus losers. *Psychology of Addictive Behaviors*, 19(3) :311–316.

Xiao, J., Ehinger, K. A., Hays, J., Torralba, A., and Oliva, A. (2014). Sun database : Exploring a large collection of scene categories. *International Journal of Computer Vision*, pages 1–20.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., and Torralba, A. (2010). SUN database : Large-scale scene recognition from abbey to zoo. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3485–3492.

Xue, S.-F., Tang, H., Tretter, D., Lin, Q., and Allebach, J. (2013). Feature design for aesthetic inference on photos with faces. In *2013 IEEE International Conference on Image Processing*, pages 2689–2693. IEEE.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, number 8689 in Lecture Notes in Computer Science, pages 818–833. Springer International Publishing. DOI : 10.1007/978-3-319-10590-1_-53.

Zola-Morgan, S., Squire, L. R., and Amaral, D. G. (1986). Human amnesia and the medial temporal region : enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *The Journal of Neuroscience*, 6(10) :2950–2967.

UNIVERSITE BRETAGNE LOIRE

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Romain COHENDET

Prédiction computationnelle de la mémorabilité des images : vers une intégration des informations extrinsèques et émotionnelles

Computational understanding of image memorability: towards the integration of emotional and extrinsic information

Résumé

La mémorabilité des images est un sujet de recherche récent en vision par ordinateur. Les premières tentatives ont reposé sur l'utilisation d'algorithmes d'apprentissage pour inférer le degré de mémorabilité d'une image d'un ensemble de caractéristiques de bas niveau. Dans cette thèse, nous revenons sur les fondements théoriques de la mémorabilité des images, en insistant sur les émotions véhiculées par les images, étroitement liées à leur mémorabilité. En considération de cet éclairage théorique, nous proposons d'inscrire la prédiction de la mémorabilité des images dans un cadre de travail plus large, qui embrasse les informations intrinsèques mais également extrinsèques de l'image, liées à leur contexte de présentation et aux observateurs. En conséquence, nous construisons notre propre base de données pour l'étude de la mémorabilité des images ; elle sera utile pour éprouver les modèles existants, entraînés sur l'unique vérité terrain disponible jusqu'alors. Nous introduisons ensuite l'apprentissage profond pour la prédiction de la mémorabilité des images : notre modèle obtient les meilleures performances de prédiction à ce jour. En vue d'amender ces prédictions, nous cherchons alors à modéliser les effets contextuels et individuels sur la mémorabilité des images. Dans une dernière partie, nous évaluons la performance de modèles computationnels d'attention visuelle, de plus en plus utilisés pour la prédiction de la mémorabilité, pour des images dont le degré de mémorabilité et l'information émotionnelle varient. Nous présentons finalement le film interactif « émotionnel », qui nous permet d'étudier les liens entre émotion et attention visuelle dans les vidéos.

Mots clés

Prédiction de la mémorabilité des images, Émotion, Apprentissage profond, Attention visuelle, Idiosyncrasie

Abstract

The study of image memorability in computer science is a recent topic. First attempts were based on learning algorithms, used to infer the extent to which a picture is memorable from a set of low-level visual features. In this dissertation, we first investigate theoretical foundations of image memorability; we especially focus on the emotions the images convey, closely related to their memorability. In this light, we propose to widen the scope of image memorability prediction, to incorporate not only intrinsic, but also extrinsic image information, related to their context of presentation and to the observers. Accordingly, we build a new database for the study of image memorability; this database will be useful to test the existing models, trained on the unique database available so far. We then introduce deep learning for image memorability prediction: our model obtains the best performance to date. To improve its prediction accuracy, we try to model contextual and individual influences on image memorability. In the final part, we test the performance of computational models of visual attention, that attract growing interest for memorability prediction, for images which vary according to their degree of memorability and the emotion they convey. Finally, we present the "emotional" interactive movie, which enable us to study the links between emotion and visual attention for videos.

Key Words

Image memorability prediction, Emotion, Deep learning, Visual attention, Idiosyncrasy