

# Contribution des caractéristiques diagnostiques dans la reconnaissance des expressions faciales émotionnelles : une approche neurocognitive alliant oculométrie et électroencéphalographie

Thèse de doctorat de l'Université Paris-Saclay  
préparée à Laboratoire Complexité, Innovation, Activités Motrices et  
Sportives (CIAMS) et  
à Laboratoire de recherches cliniques et en santé publique sur les  
handicaps psychiques, cognitifs et moteurs (HANDIRéSP)

École doctorale n°566 Dénomination et sigle  
Spécialité de doctorat : Psychologie

Thèse présentée et soutenue à Orsay de soutenance, le 11/5/2018, par

**Yu-Fang Yang**

Composition du Jury :

Laurence Conty Professeur, Université Paris Nanterre	Président
Anne Giersch Directrice de Recherche, INSERM, Strasbourg	Rapporteur
Matthias Gamer Professeur, University of Würzburg	Rapporteur
Sylvain Chevallier Maître de conférences, Université de Versailles Saint-Quentin-en-Yvelines	Examineur
Laurence Conty Professeur, Université Paris Nanterre	Examineur
Michel-Ange Amorim Professeur, Université Paris-Sud	Directeur de thèse
Eric Brunet-Gouet Praticien-Hospitalier, Université de Versailles Saint-Quentin-en-Yvelines	Co-Directeur de thèse

**Titre :** Contribution des caractéristiques diagnostiques dans la reconnaissance des expressions faciales émotionnelles : une approche neurocognitive alliant oculométrie et électroencéphalographie

**Mots clés :** Emotion, Traitement du visage, Potentiels évoqués (ERPs), Eye-tracking, schizophrénie

**Résumé :** La reconnaissance experte de l'expression faciale est cruciale pour l'interaction et la communication sociale. Le comportement, les potentiels évoqués (ERP), et les techniques d'oculométrie peuvent être utilisés pour étudier les mécanismes cérébraux qui participent au traitement visuel automatique. La reconnaissance d'expressions faciales implique non seulement l'extraction d'informations à partir de caractéristiques faciales diagnostiques, stratégie qualifiée de traitement local, mais aussi l'intégration d'informations globales impliquant des traitements configuratifs. Des nombreuses recherches concernant le traitement des informations faciales émotionnelles il apparaît que l'interaction des traitements locaux et configuratifs pour la reconnaissance des émotions est mal comprise. La complexité inhérente à l'intégration de l'information faciale est mise en lumière lorsque l'on compare la performance de sujets sains et d'individus atteints de schizophrénie, car ces derniers ont tendance à s'attarder sur quelques éléments locaux, parfois peu informatifs. Les différentes façons d'examiner les visages peuvent avoir un impact sur la capacité socio-cognitive de reconnaître les émotions. Pour ces raisons, cette thèse étudie le rôle des caractéristiques diagnostiques et configuratives dans la reconnaissance de l'expression faciale. En plus des aspects comportementaux, nous avons donc examiné la dynamique spatiale et temporelle des fixations à l'aide de mesures oculométriques, ainsi que l'activité électrophysiologique précoce considérant plus particulièrement les composantes P100 et N170. Nous avons créé de nouveaux stimuli des esquisses par une transformation numérique de portraits photos en esquisses, pour des visages exprimant colère, tristesse, peur, joie ou neutralité, issus de la base Radboud Faces Database, en supprimant les informations de texture du visage et ne conservant que les caractéristiques diagnostiques (yeux et sourcils, nez, bouche).

Ces esquisses altèrent le traitement configural en comparaison avec les visages photographiques, ce qui augmente le traitement des caractéristiques diagnostiques par traitement élémentaire, en contrepartie. La comparaison directe des mesures neurocognitives entre les esquisses et les visages photographiques exprimant des émotions de base n'a jamais été testée, à notre connaissance. Dans cette thèse, nous avons examiné (i) les fixations oculaires en fonction du type de stimulus, (ii) la réponse électrique aux manipulations expérimentales telles que l'inversion et la déconfiguration du visage.

Concernant, les résultats comportementaux montrent que les esquisses de visage transmettent suffisamment d'information expressive (compte tenu de la présence des caractéristiques diagnostiques) pour la reconnaissance des émotions en comparaison des visages photographiques. Notons que, comme attendu, il y avait un net avantage de la reconnaissance des émotions pour les expressions heureuses par rapport aux autres émotions. En revanche, reconnaître des visages tristes et en colère était plus difficile. Ayant analysé séparément les fixations successives, les résultats indiquent que les participants ont adopté un traitement plus local des visages croqués et photographiés lors de la deuxième fixation. Néanmoins, l'extraction de l'information des yeux est nécessaire lorsque l'expression transmet des informations émotionnelles plus complexes et lorsque les stimuli sont simplifiés comme dans les esquisses. Les résultats de l'électroencéphalographie suggèrent également que les esquisses ont engendré plus de traitement basé sur les parties. Les éléments transmis par les traits diagnostiques pourraient avoir fait l'objet d'un traitement précoce, probablement dû à des informations de bas niveau durant la fenêtre temporelle de la P100, suivi d'un décodage ultérieur de la structure faciale dans la fenêtre temporelle de la N170. En conclusion, cette thèse a permis de clarifier certains éléments de la

discussion concernant le traitement des visages par configuration et par partie pour la reconnaissance des émotions, et d'approfondir notre compréhension actuelle du rôle des caractéristiques diagnostiques et des informations configurales dans le traitement neurocognitif des expressions faciales des émotions.



**Title:** The contribution of diagnostic featural information to the recognition of emotion facial expressions: a neurocognitive approach with eye-tracking and electroencephalography

**Keywords:** Emotion, Face processing, Event-related potentials (ERPs), Eye-tracking, schizophrenia

**Abstract :** Proficient recognition of facial expression is crucial for social interaction. Behaviour, event-related potentials (ERPs), and eye-tracking techniques can be used to investigate the underlying brain mechanisms supporting this seemingly effortless processing of facial expression. Facial expression recognition involves not only the extraction of expressive information from diagnostic facial features, known as part-based processing, but also the integration of featural information, known as configural processing. Despite the critical role of diagnostic features in emotion recognition and extensive research in this area, it is still not known how the brain decodes configural information in terms of emotion recognition. Complexity of facial information integration becomes evident when comparing performance between healthy subjects and individuals with schizophrenia because those patients tend to process featural information on emotional faces. The different ways in examining faces possibly impact on social-cognitive ability in recognizing emotions. Therefore, this thesis investigates the role of diagnostic features and face configuration in the recognition of facial expression. In addition to behavior, we examined both the spatiotemporal dynamics of fixations using eye-tracking, and early neurocognitive sensitivity to face as indexed by the P100 and N170 ERP components.

In order to address the questions, we built a new set of sketch face stimuli by transforming photographed faces from the Radboud Faces Database through the removal of facial texture and retaining only the diagnostic features (e.g., eyes, nose, mouth) with neutral and four facial expressions - anger, sadness, fear, happiness. Sketch faces supposedly impair configural processing in comparison with photographed faces, resulting in increased sensitivity to diagnostic features through part-based processing. The direct comparison of

neurocognitive measures between sketch and photographed faces expressing basic emotions has never been tested. In this thesis, we examined (i) eye fixations as a function of stimulus type, and (ii) neuroelectric response to experimental manipulations such face inversion and deconfiguration. The use of these methods aimed to reveal which face processing drives emotion recognition and to establish neurocognitive markers of emotional sketch and photographed faces processing.

Overall, the behavioral results showed that sketch faces convey sufficient expressive information (content of diagnostic features) as in photographed faces for emotion recognition. There was a clear emotion recognition advantage for happy expressions as compared to other emotions. In contrast, recognizing sad and angry faces was more difficult. Concomitantly, results of eye-tracking showed that participants employed more part-based processing on sketch and photographed faces during second fixation. The extracting information from the eyes is needed when the expression conveys more complex emotional information and when stimuli are impoverished (e.g., sketch). Using electroencephalographic (EEG), the P100 and N170 components are used to study the effect of stimulus type (sketch, photographed), orientation (inverted, upright), and de-configuration, and possible interactions. Results also suggest that sketch faces evoked more part-based processing. The cues conveyed by diagnostic features might have been subjected to early processing, likely driven by low-level information during P100 time window, followed by a later decoding of facial structure and its emotional content in the N170 time window. In sum, this thesis helped elucidate elements of the debate about configural and part-based face processing for emotion recognition, and extend our current understanding of the role of diagnostic features and configural information during



---

## Acknowledgments

This work would not have been possible without the financial support of The Taiwan Paris-Sud Doctoral Scholarships. I am especially indebted to my supervisors Michel-Ange Amorim and Eric Bunet, who have been supportive and who worked actively to provide their guidance and feedback support for my Ph.D. study.

I also grateful to my thesis committee: Prof. Mattias Gamer, Prof. Anne Giersch, Prof. Laurence Conty, and Dr. Sylvain Chevallier, for their insightful comments and encouragement, but also for the hard question which incited me to widen my research from various perspectives.

I would like to thank Prof. Christine Le Scanff for her great support and her excellent yoga classes. Also a big thank you to Marie-Pierre for her invaluable assistance and support.

I would like to extend my gratitude to all of those with whom I have had the pleasure to work during my Ph.D. study and other related projects. Thanks, Elise for her patience and guidance, it was a memorable experience to work together. Also thanks Dorian, Mariana, Emmanuel, Sylvain Chevallier and Jean for their insightful comments and suggestions which have been valuable additions to this thesis.

I would also like to thank my colleagues for creating countless delightful memories and sharing rich French culture (good/bad side) with me. I am grateful for their friendship, inspiration and continued interest in my work. Thanks, Lea, Florie, Hugo, Mäelle, Maxime, Stephane, Michael, Elodie & Elodie, Servane, Ariane, Bing-Bing, and Hoan for their warm welcome, valuable support. A big thank to Anne, Irene, Aurore, Jie-Wen, Guillaume, Elodie for their valuable hints and comments to help my work. A very special thank you goes to Nicolas & Jean who fix all the problem related to soft/hard-ware. And Nicolas, Agnès, Thomas, Bastien, Carole, Anelore, Paul, Isabelle, Mathieu for their good advice and lovely conversation. Also thanks to Jean-Michel, and Do for their support. And thanks Adel for his help.

The support of close friends deserves a special acknowledgment - Zhou-Jian, Sajad, Szu-Han, Doris, Cam, Lin, Acer, Shashi, Hsin-Ping, Min-I, Chen-You, Ying-Chu, & Eric - fun times with you guys provided some much-needed balance and thank you guys always being there!

In particular, I would like to thank the participants (anonymous in this thesis :), my colleges, and friends - their sacrifice of time and energy to take part in my experiments (particular EEG studies) has been the main factor that has made so much of this thesis possible!

Nobody has been more important to me in the pursuit of this Ph.D. study than my family and Wegner family. They are the ultimate role for their encouragement and understanding. Most importantly, I wish to express profound appreciation to David who has been supportive, loving, critique and patience during these hard time of the study.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Literature review and research questions</b>	<b>7</b>
2.1	Facial expression processing . . . . .	7
2.2	The mechanism of face processing . . . . .	15
2.3	The role of configural information in emotion recognition . . . . .	20
2.4	Emotion recognition . . . . .	26
2.4.1	Behavioral measures . . . . .	26
2.4.2	Eye-tracking approach . . . . .	28
2.4.3	Electroencephalographic (EEG) approach . . . . .	38
2.5	Hypotheses . . . . .	46
<b>3</b>	<b>Experimental results</b>	<b>51</b>
3.1	Validation of sketch faces . . . . .	52
3.1.1	Hypotheses . . . . .	52
3.1.2	Methods . . . . .	52
3.1.3	Results . . . . .	60
3.1.4	Discussion . . . . .	62
3.2	Fixation patterns of sketch and photographed faces . . . . .	64
3.2.1	Hypotheses . . . . .	64
3.2.2	Discussion . . . . .	65
3.3	<i>Publication 1</i> . . . . .	68
3.4	ERP responses on face inversion and deconfiguration . . . . .	112
3.4.1	Preliminary EEG data inspection . . . . .	112
3.4.2	Hypothesis Face inversion . . . . .	118
3.4.3	Hypothesis Face deconfiguration . . . . .	118
3.5	<i>Publication 2</i> . . . . .	120
<b>4</b>	<b>General discussion</b>	<b>161</b>
4.1	Behavioral performance . . . . .	162
4.1.1	Sketch faces . . . . .	162
4.1.2	Comparison between sketch and photographed faces . . . . .	164
4.2	Fixation patterns . . . . .	166
4.2.1	The mean fixation duration and location . . . . .	166
4.2.2	The first and second fixation duration and location . . . . .	167
4.2.3	Summary . . . . .	168
4.3	Influence of sketch and photographed face stimuli on ERPs . . . . .	169
4.3.1	P100. . . . .	170

---

4.3.2	N170 . . . . .	171
4.4	Outlook and perspective . . . . .	174
4.5	Conclusion . . . . .	179
<b>5</b>	<b>Supplementary data</b>	<b>181</b>
5.1	EEG pilot study on individuals with schizophrenia . . . . .	181
5.1.1	Hypotheses . . . . .	181
5.1.2	Method . . . . .	182
5.1.3	Analyses . . . . .	184
5.1.4	Summary of results . . . . .	184
5.1.5	Conclusion and discussion . . . . .	188
	<b>Bibliography</b>	<b>212</b>
<b>6</b>	<b>Résumé/Abstract</b>	<b>213</b>

# Introduction

In daily social life, humans continuously decipher emotional cues with thousands of eye fixations on hundreds of diverse face information while encountering people on different occasions. The ability to recognize facial expressions rapidly and accurately is essential in order to respond appropriately. The preferential response of infants to simple face-like patterns show that we explore faces already from a very young age ([De Haan et al., 2002](#)). In general, humans show a preferential and significantly longer fixation on faces than any other object within an observed scene or an image ([Haxby et al., 2000](#)). Interestingly the length of the fixation is influenced by the facial expression itself ([Guo et al., 2006](#)). The eye movements never rest entirely, and the uptake of information occurs rapidly between the fixations ([Martinez-Conde et al., 2004](#)). Thereby, the visual system is either voluntarily engaged in an active exploration of facial expression stimuli or eye movements are automatically driven (attracted) by salient features/cues ([Langner et al., 2010](#); [Scheller et al., 2012](#)).

The two central mechanisms of face processing are part-based processing and configural processing ([Maurer et al., 2002](#); [Rossion et al., 2010](#); [Bruce and Young, 2012](#); [Meaux and Vuilleumier, 2016](#)). The former one refers to the extraction of information from one part of the face, whereas the latter one implies the integration of all facial features into a united representation ([Piepers and Robbins, 2012](#); [Rossion and Gauthier, 2002](#); [Sagiv and Bentin, 2001](#); [Leder and Bruce, 2000](#); [Tanaka and Simonyi, 2016](#)). Human observers deploy those two processing strategies interchangeably depending on the complexity and purpose of the presented facial stimuli ([Meaux and](#)

Vuilleumier, 2016). However, these processes can also be impaired for people with specific neurodevelopmental disorders. For instance, individuals with schizophrenia or with autistic spectrum disorder have a preference for processing local featural information (e.g., part-based processing) whereas healthy subjects are more likely to integrate the facial information (i.e., configural processing) (Clark et al., 2013; Bediou et al., 2005; Spezio et al., 2007; Baron-Cohen et al., 1995). This difference in visual processing is leading to an impairment in recognizing emotions correctly and can be related to an inability to obtain information from the specific region of a face - also termed diagnostic facial features (e.g., the eyes, nose, mouth).

In numerous studies, the role of diagnostic features for the perception of facial expressions has been widely acknowledged, but very few of them have compared with the role of configural processing. In order to investigate the importance of face configural information, a crucial factor is the used of face stimuli. Apart from photographed faces other stimuli were used to demonstrate the importance of configural information - such as: inverted faces (Gauthier et al., 2000; Rossion and Boremanse, 2008; Yin, 1969), scrambled faces (Bombari et al., 2013; Tanaka and Farah, 1993) or spatial frequencies filtered faces (Goffaux and Rossion, 2006; Schyns and Oliva, 1999). Recent studies found that sketch faces have shown to influence the accessibility of configural information on face stimuli (Meinhardt-Injac et al., 2013; Tzschaschel et al., 2014; Zhao et al., 2016). Sketch faces depict the face information into edges and shapes as a way of representing information from photographed faces, but the lack of facial texture (e.g., color, shadow, etc.) appears to reduce configural processing, resulting in part-based processing (Meinhardt-Injac et al., 2013; Zhao et al., 2016). To our knowledge, no study had used sketch faces and directly compared them with their photographed face counterparts to address the question of emotional face processing. To understand better the role of diagnostic features in the context of full-face configuration, the comparison of sketch and photographed faces might shed light on how human observers decode facial expressions.

One method to examine the spatial attention and temporal aspects of facial expression decoding strategies is by monitoring the eye movement using Eye-tracking. The diagnostic facial features (the eyes, nose, mouth) contribute significantly to suc-

cessful facial expression recognition. Several studies have demonstrated that the eye and mouth regions are the most diagnostic for the emotion recognition ([Gosselin and Schyns, 2001](#); [Smith et al., 2005](#)). Therefore, numerous studies divide the emotional face into “upper” and “lower” face parts for the analysis of the preferential regions ([Eisenbarth and Alpers, 2011](#); [Scheller et al., 2012](#)). However, facial expressions might not always be recognized merely by a single facial feature especially when the number of facial expressions increases within the study ([Bombari et al., 2013](#); [Smith et al., 2005](#)). For instance, participants tend to misclassify an angry expression as a neutral expression and sadness as a disgusted expression when the number of facial stimuli increases in an emotion recognition task ([Palermo and Coltheart, 2004](#)). A potential explanation could be that those facial expressions contain similar relevant information in the diagnostic features such as in the eye region ([Fiorentini and Viviani, 2009](#)), or the nose region in the case of the fearful faces and disgusted faces ([Bombari et al., 2013](#); [Jack et al., 2009](#)). Recent studies suggest that facial emotion recognition involves both processing of featural information (i.e., part-based processing) and integration of expressive components (i.e., configural processing) ([Meaux and Vuilleumier, 2016](#); [Fiorentini and Viviani, 2009](#)). The importance of integrating facial features into a face configuration to successfully recognize emotion becomes evident when comparing the performance of healthy subjects with patients with schizophrenia or autism spectrum disorder ([Bediou et al., 2005](#); [Kohler et al., 2003](#)). Patients with schizophrenia show that the difficulty in processing configural information results in perceiving the restrictive information by focusing on featural information ([Loughland et al., 2004](#)). However, until now only a few studies have been addressed the importance of configural information in emotion recognition on healthy subjects.

Another interesting method for the investigation of emotional processing is using electroencephalography (EEG) and evaluating the event-related potentials (ERPs), such as the P100 and N170 ([Bentin et al., 1996](#); [Rossion et al., 1999](#); [Vuilleumier and Pourtois, 2007](#); [Eimer and Holmes, 2007](#)). A particular interesting ERP in the context of face perception is the N170, which reflects the activation of the occipito-temporal cortex in response to face stimuli and is observed at around 170 ms after stimulus onset. ([Eimer, 2000a](#); [Caharel et al., 2013](#); [Latinus and Taylor, 2005](#); [Rossion, 2014](#)). In general, the recognition of facial expressions relies not only heavily on featural information

(such as the eye region, nose and mouth ([Itier et al., 2007](#); [Eimer, 1998](#); [Nemrodov et al., 2014](#))), but also depends on the configural information ([Batty and Taylor, 2003](#); [Caharel et al., 2009](#)). These processing modes (part-based vs. configural) result in a observable modulation of the N170 component depending on facial expressions ([Blau et al., 2007](#); [Batty and Taylor, 2003](#); [Itier and Taylor, 2004a](#)). Nevertheless, some studies show that the face-sensitive N170 component is not affected by sketch faces ([Sagiv and Bentin, 2001](#)), suggesting the sketch face is less likely to evoke configural processing during this time window. This evidence shows that more in-depth studies are necessary to understand the reason for such contradictory findings better.

In order to address the question of how part-based and configural processing are contributing to emotion recognition, we created a new set of sketch face stimuli by transforming photographed faces from the Radboud Faces database ([Langner et al., 2010](#)) with neutral expressions and four basic emotions - happiness, anger, fear, sadness. The principal intention of this thesis is to investigate the role of facial features for emotion recognition using eye-tracking and EEG measurements. Those two techniques were applied to study similarities and differences in emotion recognition between the created sketch face stimuli and their photographed face counterparts stimuli regarding 1) spatio-temporal organization of fixation (i.e., first fixation and second fixation); 2) changes in the face-sensitive N170 component, both supported by behavioral measures. To date, there are no studies which have examined the perception of facial expressions of emotional sketch faces and their emotional photographed face counterpart to this extent. In this thesis, we will disentangle the mechanisms of configural/holistic and part-based (featural/local) processing on healthy participants by observing their visual explorations for both sketch and photographed faces. A focus is put on the investigation of the contribution of the diagnostic facial features like eyes, nose and mouth for emotional expression recognition. In this thesis, the presented studies establish fundamental baselines of healthy subjects for future experiments with people suffering from neurodevelopmental disorders. A first preliminary study of individuals with schizophrenia was also conducted and shows the difficulties in obtaining high-quality data.

In the first part of this thesis, the newly created sketch faces were validated in

an emotion recognition task with behavioral measures (Experiment 1, see 3.1). We then selected the ten best recognizable sketch face stimuli per emotion that were all above the conservative threshold ( $Hu > .75$ ) to be able to compare them with their photographed counterparts in the subsequent studies. The selected sketch faces were then used to compare with their photographed face counterparts in the second study using eye-tracking to study how the perception of emotion recognition varied with face stimulus type (sketch vs. photographed faces). Gaze analysis allows us to probe the effect of sketch and photographed stimuli on the fixation duration and location on emotional facial features in the context of intact full faces that presumably allow both configural/holistic and part-based processing to take place. It is worth acknowledging that most of the past studies either used photographs or sketch faces, whereas our experimental design allowed a direct comparison with a classical emotion recognition paradigm (Experiment 2, Original publication 1, see 3.3).

In the second part of the thesis, we evaluated the ERPs components P100 and the face sensitive N170, which were obtained by EEG measurements in response to the face processing of upright vs. inverted photographed or sketch faces. We found evidences that different neuronal activities are required for the processing of photographed faces compared to sketch faces. Emotion recognition of photographs was further affected by face inversion resulting in more part-based processing, whereas this mechanism of processing was found for upright sketch faces. These data suggest that emotion recognition can be achieved successfully from isolated facial features of sketch faces, but did not necessary trigger configural/holistic processing. Furthermore, in order to better understand face processing of the sketch faces, we manipulated the spatial distance between facial features parametrically to determine whether the N170 component reflects the neuronal recruitment related to configural information processing. Although the spatial distance of the facial configuration is distorted, we observed that facial features are still informative for emotion recognition. Participants carried out high level behavioral performance without the N170 component being affected by the distorted configuration of faces (Experiment 3 & 4, Original Publication 2, see 3.5).

In summary, this thesis is divided in five chapters. After this chapter sum-

marizing the thesis, the second chapter provides a general literature review on face expression processing, the contribution of configural and part-based information, as well as of diagnostic features, and evidences from studies using eye-tracking or EEG techniques to explore emotional face processing with their limitations. The following chapter provides the results of the four experiments that were derived from the specific research question for each part of the thesis. The first (Experiment 1 & 2, respectively in the Chapter 3.1 & 3.2) and second study (Experiment 3 & 4, respectively in the Chapter 3.4.2 & 3.4.3) are the summary of two articles in submission and in order to avoid repeated contents the details are provided in the following chapter (Publication 1 in the Chapter 3.3 and Publication 2 in the Chapter 3.5). The final chapter provides an overall discussion and outlines future perspectives. Following, the supplementary chapter summarizing the result of a pilot study with schizophrenic patients conducted under the experimental paradigms of Experiment 3 & 4.

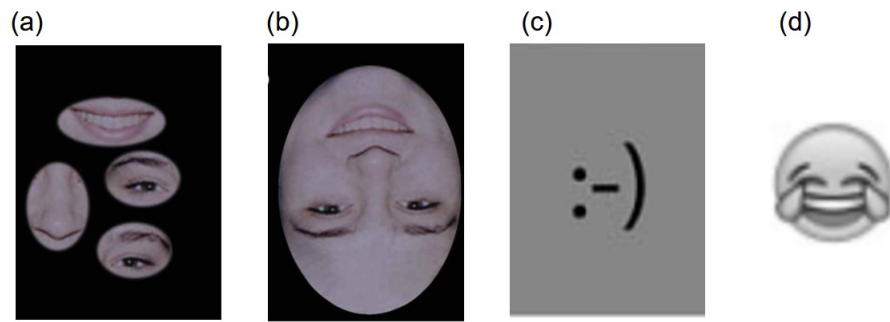
# Literature review and research questions

## 2.1 Facial expression processing

### Emotion expression

Humans are remarkably proficient in decoding emotional expressions, discriminating them and evaluating their relevance to respond to other's emotional state ([Gosselin and Schyns, 2001](#); [Smith et al., 2005](#)). A facial expression can be recognized and discriminated from a multitude of different forms of face stimuli, even with complex face stimuli such as scrambled faces ([Jacques and Rossion, 2006](#)), inverted faces ([Richler et al., 2011b](#)), line-drawing faces (e.g., smiley) ([Churches et al., 2014](#)), and equally an Emoji with tears of joys ([Skiba, 2016](#); [Aragón, 2017](#)), see Figure 2.1.

Despite the fact that the examples illustrated in Figure 2.1 show a positive expression (i.e., happiness) in different unusual and un-naturalistic configurations or with different facial textural information, human observers are able to identify of the face or even the emotional content. However, which elements are making facial expressions so unique that we can quickly identify their content in various facial configurations? Each emotion is believed to correspond to a specific internal emotional state, and can be represented as a set of facial muscle movement, termed a "facial expression" ([Gendron and Barrett, 2017](#); [Ekman and Friesen, 1978](#)). Emotions have



**Figure 2.1:** Example of different type of face stimuli for happy expressions. The images (a) & (b) were reprinted with permission from [Bombari et al. \(2013\)](#), Copyright 2013, *Quarterly Journal of Experimental Psychology*. The image (c) was reprinted with permission from [Churches et al. \(2014\)](#), Copyright 2016, *Social neuroscience*. The image (d) was reprinted with permission from [Skiba \(2016\)](#), Copyright 2016, *Nursing education perspectives*

been categorized using a set of necessary and sufficient features. This specific face configuration is essential for labeling them, and to be able to distinguish them from other emotions.

The facial muscle movements are considered to be adaptive (necessary for survival) and resulting from an evolutionary process. There is evidence that homologous facial muscle movements are shared with human and non-human animals (for review see [Waller and Micheletta, 2013](#)). This notable taxonomic approach was inspired by Charles Darwin (1916/1872) who observed that humans and non-human animals expressing stereotyped facial muscle movements, and he assumed that these patterns give the evidence of shared emotions (the original version was on 1872; [Darwin et al., 2009](#)). Darwin focused on the adaptive function of facial muscle movement for the evolutionary continuity between human and other animals. In the 1920s, [Allport and Allport \(1921\)](#) postulated that there was "functional" value in the facial muscle movements in which serve the function of different emotions in human social communication. The role of the face in the differentiation of emotion was then investigated in more detail in the 1960s and 1970s by Ekman and Tomkins ([Ekman et al., 1971](#); [Tomkins, 1962](#); [Tomkins and McCarter, 1964](#)). Thereby the term of facial "expression" implies (assumption) that a face "expresses" an individual's internal emotional state ([Langner et al., 2010](#); [Ekman, 1972](#)), but also see [Barrett et al., 2011](#) for emotion perception with body). In the early 1990s, psychological researchers used two methods: portrayal paradigm and forced-choice responses to identify the specific region(s) of

facial muscle movement that leads to the stereotype expression of emotions (e.g., [Bell, 1806](#); [Frois-Wittman, 1930](#)). The portrayal paradigm refers to the use of a posed and static expression that usually contains images of expression which are instructed to configure their face to display certain expressions. Those two methods are still commonly used to study perception of facial expressions (e.g., [Calvo and Nummenmaa, 2016](#); [Langner et al., 2010](#); [Blais et al., 2012](#)).

## Basic emotions

One of the early pioneering scientists for facial expression research is Paul Ekman. He proposed that specific facial configurations corresponding to specific emotions are universal ([Ekman, 1972](#)). Further, he claimed a ubiquitous nature of certain aspects of emotional expression in humans. For example, he found that in situations of fear, the wide-open mouth and eyes can be found in a tribe in Papua New Guinea but also in the western world or elsewhere ([Ekman, 1972](#)). For six (so-called “basic”) emotions he provided evidence suggesting a universal signaling behavior, including happiness, surprise, fear, sadness, anger, and disgust ([Ekman, 1992](#)). Still, it is still under debate whether these six basic expressions are universal ([Jack et al., 2012](#)). This concept of universal signaling behavior was initially proposed by Darwin explaining that the natural selection resulted in the development of emotion as a universal, cross-cultural experience ([Darwin, 1998](#)).

As the expression of emotions plays a large role in our daily life, categorizing them objectively was the first challenge for researchers. Therefore, some observer-based systems for facial expression measurements have been developed ([Izard, 1992](#); [Cohn et al., 2007](#); [Ekman and Friesen, 1978](#)). A popular and widely used system for the description of facial expressions is the Facial Action Coding System (FACS) ([Ekman and Friesen, 1978](#); [Ekman et al., 2002](#)), which fractionate any facial movement into anatomically based minimal “action units” (AUs) (see [Ekman and Friesen 1976, 2003](#)). These AUs correspond to specific facial movements, which are individually labeled, and can be grouped into certain combinations that account for emotion. For instance,

**Table 2.1:** Emotional facial expressions activate specific Action Units (AUs) (Langner et al., 2010)\*.

AU	AOI	AU name	Anger	Fear	Neutral	sad	happy
1	Eyes	inner brow raiser		x		x	
2		outer brow raiser		x			
4		brow lowerer	x	x		x	
5		upper lid raiser	x	x			
7		lid tightener	x				
12	Mouth	lip corner puller					x
15		lip corner depressor			x		
20		lip stretcher		x			
23		lip tightener	x				
24		lip pressor	x				
25		lip parted		x			x
6		cheek raiser					x
17		chin raiser	x			x	

*Note.* AOI = Area of interest.

\* The description of AUs has been modified according to the original text in Ekman and Friesen (1976).

a happy expression activates AU 6 (cheek raise), AU 12 (lip-corner pull), and AU 25 (lips part) (see Langner et al., 2010, also see in Table 2.1).

A requirement of FACS is the assumption that posed, static configurations of facial muscle movements provide sufficient cues to recognize emotion. Thus, FACS is widely used for validating face databases, such as Pictures of Facial Affect (POFA: Ekman and Friesen 1976); Karolinska Directed Emotional Faces database (KDEF: Lundqvist et al. 1998); Warsaw Set of Emotional Facial Expression Pictures (WSEFEP: Olszanowski et al. 2014); or the Radboud Faces Database (RaFD: Langner et al. 2010). Although those face stimuli were validated using the FACS, not all facial features are equally well distinctive for recognizing emotion. Elsherif et al. (2017) found that a smiling mouth is easier to be identified than a fearful mouth in the Ekman and Friesen's set (1975). They suggest that Ekman and Friesen's set aimed at posing prototypical facial expressions, whereas the KDEF set used spontaneous expressions posed by actors. Because the fearful mouth bares their teeth (e.g., not necessary revealing teeth), whereas the smiling mouth exposes their teeth in the Ekman and Friesen's set, but it is not the case in the KDEF set. The contrast between the white teeth and the darker lip on smiling mouth might enhance the advantage effect in detecting happy faces in the



**Figure 2.2:** Wegrzyn et al. (2017) used the main metric to assign 48 face tiles for each face, computed by formula (see their study for details). The tile got a weight from the sum across trials for each participant and expression. The larger the tile the more often it was short responded and correct trials. The weight is visualized with a green-red color bar which represents mix-max scale for each expression. Green color indicates the lowest weight, and red color shows the highest weights. The image adapted from (Wegrzyn et al., 2017), *Journal of PloS one*, with permission.

Ekman and Friesen's set. These posed face stimuli are to some extent artificial facial expressions stimuli that may not always be ideal to investigate how human perceive emotional faces in real life and vary depending on the database sets.

Additionally, the process of recognizing facial expressions is driven by low-level visual properties located in specific regions (e.g., the eyes, and mouth), see Figure 2.2, whereas AUs capture the minimal muscle movements of how to express certain emotions (i.e., artificially posed expressions). In summary, AUs provide a physical explanation of the cause of facial expression, and the facial feature represents the salient informative zone of emotional expression.

## The saliency of expressive facial features

Human observers tend to look to the most relevant features of a scene, in particular faces (Henderson et al., 2005). Thereby the majority of eye fixations is directed on the internal features, like the eyes, nose, and mouth, and not on external features such as ears or hair (Yarbus, 1967; Walker-Smith et al., 2013). Emotionally expressive faces seem

to be preferentially processed as compared to neutral faces (Palermo and Rhodes, 2007; Gamer et al., 2013). The findings support the concept that the ability to understand and interpret one's displayed emotion is a crucial communication tool for social being (Adolphs et al., 2002; Calder et al., 1996). Hanawalt (1944) showed the important role of facial features in decoding facial expressions and in distinguishing between different emotions. For example, it is suggested that the eyes convey important cues to recognize fear expressions, whereas the mouth is more informative for happy faces (Hanawalt, 1944; Schyns et al., 2009; Smith et al., 2005; Elsherif et al., 2017). An inability to spontaneously look into the eyes region impairs the ability to recognize fearful faces, as evidenced among people with cognitive disorders, such as prosopagnosia (Caldara et al., 2005), in autism (Wallace et al., 2008), schizophrenia (Morris et al., 2009; Bortolon et al., 2015), or with a psychopathic personality (Boll and Gamer, 2016). These results indicate the close relationship between diagnostic facial features and the decoding of facial expression.

Observers appear to deploy different visual exploration strategies for the evaluation of the six basic emotional expressions based on the distribution of diagnostic facial features (Smith et al., 2005). However, which parts/features ("cues") of the face are the most salient for identifying a specific emotion? In the mid-1970s, researchers started to manipulate these cues in order to identify the salient part(s) relevant for emotion recognition (Davies et al., 1977; Ellis, 1986). The early studies revealed a dominance of the eye and eye-brow combination for individual face perception (Walker-Smith et al., 2013; Walker-Smith, 1978; Davies et al., 1977). This observation was supported by subsequent studies (Schyns et al., 2009; Mehoudar et al., 2014; Wegrzyn et al., 2017; Neath-Tavares and Itier, 2016). The average scanning patterns of individuals have shown that the eyes are the most frequently fixated internal facial feature followed by nose and mouth (Mehoudar et al., 2014; Gosselin and Schyns, 2001; Sekiguchi, 2011). Nevertheless, there is still a debate in how far the task itself can influence the preference to the single feature. For instance, Schyns et al. (2002) have shown that observers focused on the eyes, mouth, and the chin in face identity tasks (i.e., neutral expression), whereas the mouth is preferred for recognizing facial expressions. Furthermore, the left side of the face around the eye region was used for gender identification (Schyns et al., 2002). However, Scheller et al. (2012) found the eyes and the mouth of

facial expressions are processed irrespective of the task demands and spatial locations. Taken together, human observers exhibit a tendency to extract information from facial features, but visual scanning strategy varies depending on the task at hand.

## Prioritization of eyes

Numerous reports suggest that observers appear to rely strongly on the information from the eyes for successfully recognizing an emotion rather than from the nose or/and mouth regions (Caldara et al., 2005; Itier et al., 2007). The bias to the eye region may be caused by the ambiguities between certain expressions that require additional processing time. For example, subjects misclassify happy or fearful faces for surprised faces (Palermo and Coltheart, 2004), mistaken isolated wide-open eyes as the expression for fear, angry or even surprise (Jack et al., 2009), or opt for angry expressions instead of disgust faces when seeing nose scrunch (Pochedly et al., 2012). Some authors suggest that the eye prioritization is due to the fact that they provide essential information as *gaze-signal* that aids cooperative and mutualistic behavior in the social communication environment (Latinus et al., 2015; George et al., 2005). In comparison to other primates' eyes, the apparent contrast of the large white sclera surrounding the darker colored iris show the gaze direction of the eyes (i.e., where the eyes are looking) (Kobayashi and Kohshima, 2001, 1997).

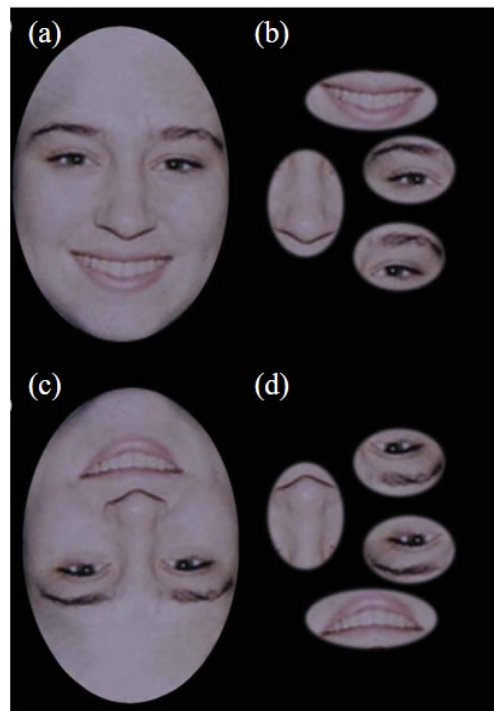
Eisenbarth and Alpers (2011) suggested that the gaze is directed to the eye region to seek for relevant emotion-specific information when the task requires participants to evaluate the valence and arousal of facial expression. According to these authors, the eye region is fixated more frequently and longer as compared to other facial regions, shown for example for recognizing sadness. However, the finding that the eye regions may be visually explored in a relatively larger proportion as compared to the nose and the mouth might be due to a methodological artifact. Because most studies are grouping the left and right eyes unitedly with their eyebrows, which increases the amount of information to be processed. There is further some inter-cultural variability in this eye prioritization behavior. For example, the Caucasian participants tend to

fixate more around the eyes region and mouth, whereas Easterners use more global information (i.e., looking at the center of the face) from the features for effective face recognition (Jacques and Rossion, 2006; Miellet et al., 2013).

## The diagnostic features

Numerous studies have shown that essential cues for the recognition of facial expressions can be found in the eyes, nose and mouth regions, known as the *diagnostic features* (Ekman and Friesen, 1969; Schyns et al., 2009; Eisenbarth and Alpers, 2011; Scheller et al., 2012). On the other hand, some studies raise the question if recognition performance of facial expressions can be improved and enhanced when treating the diagnostic features as individual components (i.e., pop-out facial feature) in the context of full-face configuration is being displayed (Schurgin et al., 2014; Vaidya et al., 2014; Bombari et al., 2013, 2009).

In order to examine the role of diagnostic features, scrambled face stimuli are widely used in which each feature (left eye, right eye, nose, and mouth) is isolated and then arranged randomly (see in the Figure 2.3). Because of the high salience of a smiling mouth for happy expressions, and of open-eyes with open-mouth for fearful faces, participants detect happy and fearful expressions better than anger and sadness from the isolated elements (Bombari et al., 2013). The lack of high salience in any single facial feature of angry and sad expressions requires further processing into face configuration, such as gathering information from the eyes, nose, and mouth (Palermo and Rhodes, 2007; Jack et al., 2009). Such visual processing demonstrates that to classify correctly complex or rich information of emotion expression, extracting information from the single facial feature is not enough. The importance of configuration has been illustrated with inverted faces. When facial features are inverted the reaction time increases, and the accuracy decreases significantly. Therefore, the manipulation of scrambled faces cannot reveal which diagnostic features provide relevant specific information to process emotional expressions when canonical configural information is not available (i.e., the eyes above the nose, they above the mouth). Also, isolated



**Figure 2.3:** Example of a happy face stimuli: (a) an intact upright face, (b) a scrambled face, (c) an inverted face, (d) an inverted scrambled face. Reprinted with permission from [Bombari et al. \(2013\)](#) , Copyright 2013, *Quarterly Journal of Experimental Psychology*.

facial stimuli are not frequently seen in daily life. The importance of high salient facial features depends on emotions ([Schurgin et al., 2014](#)). Altogether, these results suggest that the investigation on the availability of featural facial information should take the presence of configural information into account when studying the perception of emotional expression.

## 2.2 The mechanism of face processing

When we recall an emotional face or mimic how individual faces may be represented in a particular expression, there is a temptation to list separate features, for instance, "enlarge the eyes to show a surprised face" or "push mouth corners upward to express happiness". However, facial expressions can be recognized using very impoverished or unnatural face stimuli, for example, the face is depicted simply by two dots as eyes with an upward curve as mouth (e.g., a smiley [:])), objects formed as a face-like shape, such as two-tones Mooney faces ([Latinus and Taylor, 2005](#)), schematic faces ([Sagiv](#)

and Bentin, 2001; Henderson et al., 2003), or Arcimboldo face-like paintings (Caharel et al., 2013; Rossion et al., 2010). Haxby et al. (2002) suggested that processing facial identities and facial expressions share some underlying common neural mechanisms, particularly at the early stage of visual analysis. Several studies have investigated the process of face identification and gained detailed insights into its underlying processing mechanisms (Bentin and Deouell, 2000; Leder and Bruce, 2000; McKone et al., 2012; Fiorentini and Viviani, 2009). But far less is known about emotion recognition (Bombari et al., 2013).

Face identification studies manifest the involvement of two cognitive mechanisms, which are configural and part-based (or so-called, featural, local, analytic) processing. These findings suggest that the ability to recognize a face is not driven only by low-level visual properties but also relies on the configural information (Van Belle et al., 2010). The complexity of the combination of features reflects the diverse dimensions of face processing, such as age, gender, attractiveness, identification, and expression (Bruce and Young, 2012). The different facial features can be critical ingredients in the representation of faces, especially if some facial features have very distinctive characters, such as the smiling mouth of happy expression (Calvo and Nummenmaa, 2008; Ekman et al., 1987). However, in daily life, we perceive faces as a complete facial configuration rather than a series of separate facial features. Faces rely on the discrimination of exemplars in a very homogeneous category that shares highly similar global configurations where the eyes are above the nose, and the nose is above the mouth. The question arises of whether the facial expression can be identified on the basis of the local facial features within the most relevant regions, or involves necessarily the analysis of the global facial configuration.

## **Part-based processing**

Part-based processing contributes to face recognition and can be assessed independently from configural processing as defined previously by McKone (2004); Goffaux and Rossion (2006). Part-based processing depends on featural information of a face.

The discrimination of individual Faces is widely known on the basis of local facial component -known as the facial feature. The term *feature* describes detailed information about a specific face part/feature such as the color of the eye, texture or the shape of the mouth (Schwaninger et al., 2006). Piepers and Robbins (2012) argued that *features* in featural processing are referred to *emergent features* - comparable to a square, which is an enclosed area formed by four lines of equal length, and none of these lines can be processed on their own. It explains why some illusion face stimuli or object-formed face stimuli (e.g., houses) can be recognized as a face. The existing literature suggests that when referring to *featural processing* the information is relatively isolated being focused on (Diamond and Carey, 1986). Piepers and Robbins (2012) suggested that the processing of features should be referred to as “part-based processing” rather than “featural processing” when the features are not displayed isolated in the representation of the facial configuration. Therefore, part-based processing is useful as it allows to study the role of a facial feature in the context of full-face configuration.

## Configural face processing

When one considers the recognition of a face, many studies agree that the concept of processing the face configuration. (Leder, 1996a; Calder et al., 2000; Wang et al., 2015; Adolphs, 2002). There are two notions of processing face configuration - configural processing and holistic processing. Some researchers have separated configural processing from holistic processing (Rossion, 2008; Piepers and Robbins, 2012; Calder and Jansen, 2005), because they are sometimes equated in face studies, as they rely on a simultaneous integration of all the available information (McKone and Yovel, 2009). Holistic processing implies the integration of all visually perceived facial features at once and as a series of (memorized) templates, each of which is a single indecomposable component (Tanaka and Farah, 1993; Richler et al., 2008). However, it is not necessarily the case for configural processing where the spatial organization of the facial features mainly matters (Bartlett et al., 2003; Tanaka and Farah, 1993). Therefore, configural processing requires a spatial analysis of parts and their relationships, whereas holistic processing matches a face as a typical canonical arrangement. A major problem with

adopting holistic processing to recognize facial expressions is that a high memory load would be required for matching a multitude of templates, with one template per change in facial features and combination of changes (Piepers and Robbins, 2012). In contrast, facial expressions can be easier recognized through configural processing of spatial information carried by the face, such as the variant or invariant distance between facial features or face parts (Rhodes, 2013). However, these processes are contrary to part-based processing which relies on the detection of facial expression from isolated parts the face (i.e., the eyes, nose, and mouth) without any processing of the relation between these parts (Calder et al., 2000).

Therefore, both concepts of holistic processing and configural processing should be defined cautiously when studying facial expressions. Holistic refers to the fact that visual stimuli (faces) are coded as *Gestalten* [German plural for a *Gestalt*, meaning *figure*], which are not analyzed or represented as separate facial features. Therefore, holistic processing implies that any feature combination creates a unique configuration that cannot be predicted from its single component (i.e., feature) - reflects a way of representing and processing the face stimulus. More recently, Tanaka and Simonyi (2016) clarified that the advantage of the holistic processing of a face was specific to normal and intact faces, but not to inverted, scrambled, nor non-face objects. Indeed, it is reasonable to assume that human observers are capable of perceiving the face as a gestalt in a global scale.

The recognition of emotions appears to rely on a particular scale of analysis on the facial movements to the metric distance between facial features where meaningful information emerge (e.g., the nose to mouth distance). One popular approach to study holistic face processing is to filter the local information of facial features with a low-spatial-frequency (Goffaux and Rossion, 2006) or test face stimuli while displaying them in the subject's peripheral vision (McKone, 2004). Provided that the holistic processing relies on the "face template matching" (eyes above the nose, nose above mouth), hence, face recognition is less unaffected by the blurring information of facial features. According to this framework, the internal representation of global structures of a face remains without precise edge information of facial features (Goffaux and Rossion, 2006; Van Belle et al., 2010). Investigating the metric distance between facial

features are important in categorizing a facial expression, such as the metric distance between the eyes the mouth is varied for sad and happy expressions. Therefore, in this thesis configural processing is used to refer to the examination of the integration of facial features with consideration of metric distance for different facial expressions.

## **Both configural and part-based face processing**

Studies investigating configural and part-based processing for face detection and face identity are numerous ([Hsiao and Cottrell, 2008](#); [Flevaris et al., 2008](#); [Bentin et al., 2006](#); [Piepers and Robbins, 2012](#); [Calder and Jansen, 2005](#); [McKone, 2004](#)), for review see [Duchaine and Yovel, 2015](#), but very few studies examine the interaction between configural and part-based processing with regards to the perception of facial expressions ([Bombari et al., 2013](#); [Schurgin et al., 2014](#); [Vaidya et al., 2014](#)). But the obtained results suggest that part-based and configural information are processed simultaneously due to expressive features carrying different amounts of information depending on the emotional expression. On the other hand, authors such as [Calder et al. \(2000\)](#) and [Schyns et al. \(2009\)](#) suggest that information processing during emotional recognition either implies configural processing or relies on individual features.

Studies investigating configural and part-based processing for face Other authors found that the processing mode varied with emotion valence. For example, negative expressions lead to more part-based processing, whereas positive expressions facilitate a more configural processing ([Curby et al., 2012](#)). By reaction time data, [Fraser et al. \(1990\)](#) suggested that configural processing interferes with facial features in sketch intact faces. Later on, [Calder et al. \(2000\)](#) confirmed these results using a face composite paradigm. Although some researchers are beginning to address the question about how to examine the interplay of configural and part-based processing, the more precise temporal examination and using consistent facial stimuli are needed. It is still unclear how face processing is linked to the decoding strategy in facial expressions. Taken together, there is a necessity to examine the relationship between face processing

(e.g., part-based and configural processing) and facial expression processing.

## 2.3 The role of configural information in emotion recognition

In contrast to studies investigating the underlying perceptual mechanisms of face identification, research in facial expression recognition generally focuses on the communicative value of the facial features rather than the perceptual representation. There is a general debate about if the perception of facial identities and facial expressions are either processed in dissociable pathways ([Bruce and Young, 1986](#); [Hoffman and Haxby, 2000](#); [Phillips et al., 1998](#)), or recruit (to a certain extent) common brain areas for the core face processing ([Haxby et al., 2000](#); [Vuilleumier and Pourtois, 2007](#); [Meaux and Vuilleumier, 2016](#)). Numerous studies have investigated the role of configural and part-based information in face identification, but only a few focused on the perceptual processes involved in emotion recognition ([Bombardi et al., 2013](#); [Calder et al., 2000](#)). Ekman and his colleagues contributed greatly in enhancing the understanding and knowledge of the anatomy used to produce facial expressions, but the knowledge of the perceptual processes for decoding facial expressions (from the observer's perspective) remains unclear.

The ability to decode emotional facial expressions accurately, rapidly, and to discriminate between them is an essential component of social communication. Ekman (1977, 1967) suggested that six basic emotions (anger, disgust, happy, sad, fear, and surprise) can be identified across cultures. However, even for those basic emotions, some facial expressions are especially challenging for the visual cognitive system as indicated by the literature showing that subjects are most likely to confuse surprised faces as happy or fearful expressions, and misclassify angry faces as neutral or disgusted ([Palermo and Coltheart, 2004](#)). This confusion can be caused by the fact that subjects are biasing their fixations toward the eye regions where information are too similar to discriminate certain expressions ([Jack et al., 2009](#)).

Hence, it is imperative to consider both contributions of the diagnostic features and the modes of configural processing. The integration of visual information from diagnostic features becomes evident when comparing the visual processing of healthy subjects to the patients suffer from neurodevelopmental disorders affecting emotion recognition, such as schizophrenic (Bediou et al., 2005; Clark et al., 2013; Giersch et al., 2011) and autistic patients (Spezio et al., 2007; Wallace et al., 2008). There is evidence that the integration of complex facial information is more difficult for schizophrenic and autistic patients. For example, patients with autism show a preference in local perceptual processing (e.g., features) rather than holistic/configural processing or are biased towards the examination of the lower half of the face (Spezio et al., 2007). Similar results were found in patients with brain lesions in the amygdala. Due to the inability to process the information of the eye region, they show a deficiency in decoding fearful expression (Adolphs et al., 2005). The importance of face processing is not fully explored with regard to emotion recognition. Numerous studies focused on how participant visually explore the facial features, but only a few of them have framed their findings concerning the underlying face processing mechanisms.

## Neuroimage evidences

A recent functional magnetic resonance imaging (fMRI) study suggested that configural and part-based information engage dissociable neural pathways depending on emotional face configuration (Meaux and Vuilleumier, 2016). For example, congruent configural expression processes, for the case of two matching facial expressions with consistent internal facial features, activate the fusiform, the inferior occipital area and the amygdala. On the other hand, independent facial feature analysis engages the superior temporal sulcus (STS) and prefrontal areas, such as inferior frontal gyrus (IFG) or orbitofrontal cortex (OFC). Fusiform face area (FFA) and occipital cortex (OFA) are critical for extracting shape cues for face recognition, whereas STS, amygdala, OFC, and IFG, execute a broad range of emotional and social functions (see (Langner et al., 2010; Meaux and Vuilleumier, 2016; Adolphs et al., 2005)). In sum, these findings suggest that the recognition of facial expressions recruits different neural networks.

However, it is unclear which visual information of the face engage configural and part-based processing and when these processes take place during the perception of facial expression. Therefore, this thesis examined how emotion recognition is shaped by the reciprocal interaction between configural and part-based processing.

### **Approach 1: Impairment of configural information in sketch faces**

Several approaches have been used to investigate how facial expressions are decoded. A crucial element inside these approaches is the manipulation of the stimuli, which help to distinguish between configural and part-based processing. Based on this purpose, several kinds of stimuli and manipulations have been used, such as, faces that align or misalign a different/same top and bottom half of a face to create a new stimulus (Rossion, 2008; Grand et al., 2004); changing the space between features (Le Grand et al., 2001; McKone and Yovel, 2009); faces filter with spatial frequencies (Goffaux and Rossion, 2006).

However, each study emphasizes only on one aspect of information processing. Another way to examine if configural processing or featural processing are required for decoding a facial expression is to manipulate the facial feature information. Naturalistic photograph faces display a facial shape (including a perspective, shades, etc.) as well as surface information (color, texture, etc.) and are believed to activate configural processing, whereas the removal of the facial textural information result in a line-drawing or sketch information enhancing the featural information (Zhao et al., 2016). The sketch face only retains the information specific to the facial features (such as edge, and contour of the facial features as line-drawings) (Leder, 1996a). Although sketch face stimuli containing reduced configural information, the recognition of the facial expression may rely on different internal facial features, such as the eyes, nose, and mouth (Calder et al., 2000). The shape of a single feature can even induce robustly, and high-quality posed invariant emotion perception, such as the mouth of happy expressions (Calvo et al., 2010). Still, it is unclear what makes a sketch face inducing more part-based processing than configural processing as compared to photographed

faces, even if so, whether a different extraction of facial information is required. Which information is relevant to recognize facial emotion remains to be established.

Most of the studies of emotional facial expressions using photographed faces as face stimuli ([Eisenbarth and Alpers, 2011](#); [Beaudry et al., 2014](#); [Schurgin et al., 2014](#)), which might encourage participants to utilize more configural information in comparison to sketch faces. Consequently, these findings make it more difficult to understand the relation between the configural information and part-based information on the perception of facial expressions. This thesis intends that the stimulus manipulation must be combined with eye-tracking and electroencephalogram (EEG) to disentangle those processes. To reach this goal, i.e., changing the balance between the configural and featural processing, several methods were used in the literature, such as scrambled faces ([Bombari et al., 2013](#)), composite faces ([Calder et al., 2000](#); [Meaux and Vuilleumier, 2016](#)), changing the metric distance of facial features ([Le Grand et al., 2001](#)), and inverted faces ([Leder and Bruce, 2000](#)).

Recently, removing the information of facial texture was proposed to disentangle configural and part-based while keeping the full face configuration intact ([Zhao et al., 2016](#)). In their study, neutral photographed faces are considered containing more face surface and texture information and are thought to activate configural processing, while sketch faces trigger more part-based processing on extracting featural information. Other studies found that the identification of a face is more accurate from photographs than sketch faces because sketches reduce depth cues provided by facial texture and shadows ([Liversedge et al., 1998](#); [Tzschaschel et al., 2014](#)).

[Benson and Perrett \(1991\)](#) suggested that impoverished sketch face stimuli from photographed face stimuli improve the representation of faces in some circumstances although they contain no texture, color, and shadows. They concluded that sketch faces might contain all the necessary featural information to access the properties of facial features without attending to the pictorial level (i.e., color, texture). Several studies found that the participants perform better using real faces (photograph and grayscale) than sketch face (line-drawing) in face identification task ([Tzschaschel et al.,](#)

2014). However, some behavioral studies demonstrated that participants identified more quickly and accurately on sketch face than photographed natural faces when the task requires identifying differences in local facial features, such as the composite face task (Zhao et al., 2016); the feature change detection task (Meinhardt-Injac et al., 2013). In order to disassociate the role of part-based and configural processing, using sketch face stimuli might gain more insight into the importance of face processing for the recognition of facial expressions.

## **Approach 2: Disruption of configural processing - Face inversion**

Another approach allows studying configural processing without altering the spatial face configuration namely the use of an inverted face (Yin, 1969). It has been widely documented that the inversion of faces has a much larger effect in comparison to non-face objects due to the disruption of configural information, which makes it to a preferred tool in face processing studies (Richler et al., 2011b; Rossion, 2009; Leder and Bruce, 2000; Hinojosa et al., 2015; Yin, 1970). The visual processing of configural processing becomes evident when comparing healthy subjects to individuals with congenital prosopagnosia who are impaired at face recognition and do not exhibit typical advantage for processing upright faces over inverted faces (Behrmann and Avidan, 2005).

This face inversion effect is a evidence demonstrating that faces are processed by specific brain mechanisms, and the link between configural and face processing is critically connected. The face inversion effect is very robust in a variety of conditions, such as the task demanded distinguishing between familiar and unfamiliar faces (Collishaw and Hole, 2002; Rossion and Gauthier, 2002; Hole, 1994), and matching tasks (Freire et al., 2000; Russell et al., 2007). A common finding using the face inversion effect is the hampering of configural processing, whereas the part-based processing is less affected (Valentine, 1988; Le Grand et al., 2001; Richler et al., 2011c; Namdar et al., 2015). Such disruption of configural processing results in lower recognition accuracy and increased reaction time (Rossion, 2009; Maurer et al., 2002). The difficulty

of processing inverted faces is not limited to real photographed face stimuli. Some studies also found similar results for gray face stimuli in a familiar face task as well as for (Freire et al., 2000; Russell et al., 2007) sketch face stimuli in a face detection task (Eng et al., 2017; Sagiv and Bentin, 2001).

Although the processing of a face is known to be sensitive its orientation, only a few studies investigate the contribution of configural and part-based information when categorizing inverted emotional facial expressions (Bombari et al., 2013; Derntl et al., 2009; Prkachin, 2003). These studies found a decreased accuracy when recognizing inverted facial expressions, suggesting that configural information contributes to emotion recognition. However, the results are somewhat heterogeneous with respect to the number of facial expressions are used, leading to different inversion effects for specific expressions. Due to the fact that facial features may also be analyzed separately from configural information when judging an emotion (Beaudry et al., 2014; ?), it is necessary to take the role of facial components into account. Therefore, the face inversion effect might found that (i) part-based processing is less impaired compared to configural information; (ii) participants might then benefit from the information embedded within the diagnostic feature to categorize facial expression by extracting information from the eyes or mouth region.

A popular example of face inversion, which demonstrates how configural information modulates the perception of a facial expression is the Thatcher illusion (Thompson, 1980). Participants are able to precisely point out which facial feature is rotated when the face is presented upright (see Figure 2.4 for example). The main reason is that participants tend to search for distinctive features rather than global configural information when viewing inverted faces (Latinus and Taylor, 2006; Kimchi and Amishav, 2010). When emotional faces are presented upside down, participants continue detecting the inverted face stimuli as a face but applying a different visual strategy for the analysis of the features to recognize its emotional content. For example, they search for the mouth when seeing an inverted happy face. On the other hand, participants spend more time on the nose region as a way to process configural information when judging inverted sad expressions (Bombari et al., 2013). These findings led to the debate of whether we are adopting part-based processing or



**Figure 2.4:** Example of Thatcherized face of R. Le Grand illustrated the difficulties in identifying which inverted face stimuli has been Thatcherized. Presumably, the inversion effect impaired the sensitivity to relational facial feature information. Reprinted from [Maurer et al. \(2002\)](#), *Trends in Cognitive Sciences*, with permission.

configural processing for recognizing facial expressions. However, inverted faces increase the attention toward diagnostic features, which are critical for perceiving facial expressions. If an inverted face disrupts the configural processing, leading participants to extract information from the diagnostic features, then facial features with sketch faces (portraying the features' contours) may, in turn, aiding emotion recognition from inverted faces.

## 2.4 Emotion recognition

### 2.4.1 Behavioral measures

Emotion recognition is a complex process, but being able to quickly "read-out" the information of a facial expression is pivotal for social interaction and even survival ([Darwin, 1998](#)). Numerous studies have investigated the speed (i.e., reaction time) and the accuracy of recognition for basic facial expressions like fear, disgust, sadness, happiness, anger, and surprise ([Ekman and Friesen, 1976](#); [Langner et al., 2010](#); [Elfenbein and Ambady, 2003](#)). In an emotion recognition task, participants are normally asked to categorize the displayed face stimuli on the basis of their emotional context (e.g., happy, fear, or sad) ([Langner et al., 2010](#); [Goeleven et al., 2008](#)) or emotional valence (i.e., positive, negative or neutral) ([Goeleven et al., 2008](#); [Eisenbarth and Alpers, 2011](#)). In

general, the emotion recognition performance varies with the type of facial expressions. However, numerous studies have demonstrated that the happy expression is recognized faster and more accurately than other emotions in visual search paradigms (Leppänen et al., 2004; Öhman et al., 2001; Calvo and Lundqvist, 2008) and emotion categorized paradigms (Blais et al., 2012; Eisenbarth and Alpers, 2011). This happy face advantage is not only observed for photographed faces and sketch stimuli (Eisenbarth and Alpers, 2011; Horstmann and Bauland, 2006; Leppänen et al., 2004), but also when faces were presented in the visual periphery field and an inverted orientation (Calvo and Nummenmaa, 2016).

This advantage is attributed to the processing of a highly salient and distinctive facial feature - most likely in the mouth region. Recognizing happy depends not merely on its open or closed mouth, but rather on the shape of the mouth. Calvo and Marrero (2009) proposed that the happy mouth is a *distinctive and diagnostic* feature facilitating an efficient distinction from other emotions. They assessed 16 action units (Ekman and Friesen, 2003) across six basic emotions and found that the happy mouth feature was not shared by any other emotion, see AU table 2.1. However, the perception of facial expressions does not take place instantaneously as a signal salient featured process (Adolphs, 2006). Calvo et al. (2010) suggest that the analysis of salient features is the first stage of a quick detection before retrieving meaningful expression information through a process of configural processing. Indeed, participants tend to misclassify a happy expression as a surprise expression when the eye region is covered (Fox et al., 2000), suggesting the diagnostic feature facilitate the processing speed but the global face configuration is relevant to distinguish between emotions.

Based on the fact that not all basic expressions have a *distinctive and diagnostic* feature like the happy expression, complex facial information need to be integrated in order to achieve successful emotion recognition. On the other hand, people who have shown to have difficulties in recognizing fearful faces after a brain injury could improve their recognition performance when they were instructed to pay more attention to the eye region. (Adolphs et al., 2005; Whalen, 2004). This evidence demonstrated that the wide-open eyes might be a key character for the fearful expression. However, correct fearful face categorization cannot solely achieve based on this featural information

(Blau et al., 2007). Bombari et al. (2013) demonstrated that the fearful face recognition is better when the configuration information is presented in an intact face version rather than in a scrambled face version. Due to the fact that the open mouth on the fearful face is shared with the happy face, the relevance of the eyes and mouth is important for recognizing fearful faces.

The shared action units between emotions reflect the complexity of facial information in decoding emotions and shows that it is necessary to take configural processing into account. Many studies have investigated the link of diagnostic featural information to emotion recognition performance on individual expressions (Goeleven et al., 2008; Bombari et al., 2013; Langner et al., 2010; Calvo et al., 2014). However, neglecting the effect of configural information have led to heterogeneous results of the studied behavioral performance on individual expressions. For instance, angry and sad expressions are scored less accurate than fearful and happy faces in the study of Bombari et al., 2013 which take configural processing into account, whereas the higher accuracy was found for categorizing sad faces than fearful faces in the study of Calvo et al., 2014 which focus on the comparison between facial features. It is crucial to understand how do we decode facial expressions from the behavioral level based on the different ways of examining faces. Furthermore, it will help for a better understanding of the differences of between healthy subjects and individuals with autism, schizophrenia or depression considering these patients are significantly less accurate in decoding facial expressions (Bediou et al., 2005; Bortolon et al., 2015; Wallace et al., 2008).

## 2.4.2 Eye-tracking approach

Behavioral measures provide information such as the length of the response time and the accuracy during face recognition, but can hardly be used for the explanation of the underlying neural mechanism. Faces convey a considerable amount of information such as gender, age, identity, and emotion, with the aid of eye-tracking we can investigate how attention is deployed to select relevant information and shed light

on facial expression processing. Eye-tracking has been intensively used to explore cognitive and perceptual processing for decades, such as in reading, scene perception, or face scanning (e.g., [Liversedge et al. 1998](#); [Holmqvist et al. 2011](#); [Henderson et al. 2005](#)). We continually make eye movements to gather information when reading, looking at a scene, and searching for objects. There are two essential components of eye movements studied in various tasks, namely saccades and fixations.

The saccade is a rapid eye movement with velocity as high as 500°(of visual angle) per second and typically take 30-80 ms to complete. Because the saccadic eye movement is moving so fast in order to relocate the point of fixation, only blur information is obtained during most of the saccades ([Rayner, 2009](#); [Kowler and Steinman, 1979](#)). Fixations, on the other hand, refer to the period of time when the eyes remain fairly still for about 200-300 ms. The algorithms for detecting fixation vary with commercial recording software. Fixation location and fixation duration are tightly linked to time-locked perceptual and cognitive processes. Different fixation duration and locations are determined by the task at hand and where the particular information is acquired ([Henderson et al., 2003](#)). The fixation patterns can be explained by image properties such as contrast or salience. For example, in face stimuli, participants fixate mostly the eyes, nose, and mouth, with nearly 70 % of these fixations oriented to the eye region ([Walker-Smith et al., 2013](#); [Shipley and Kellman, 2001](#)). The function of fixation at a specific facial area is to decode a given detail. Therefore, eye fixations may index where the attention is directed to locations in space.

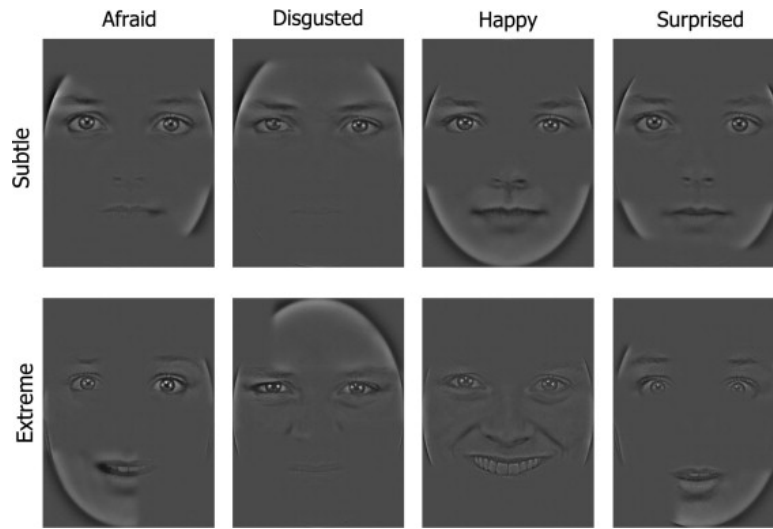
## **The preferential diagnostic features**

Eye-tracking has a long history in the investigation of attention to specific facial features ([Yarbus, 1967](#)). Fixation patterns obtained by eye-tracking technology reveal which diagnostic features are used to decode facial expressions ([Neath and Itier, 2014](#); [Schurgin et al., 2014](#); [Eisenbarth and Alpers, 2011](#)). Thereby, the fixation location and duration and the evolving fixation pattern is influenced by the observed facial expression and displayed emotion. Several studies have shown that healthy subjects

have an increased number of fixations on the eye regions while viewing fearful faces, whereas both individuals with autism and schizophrenia have a reduced amount of fixations on the eye region but increased fixations on the mouth region (Spezio et al., 2007; Lee et al., 2010; Jones et al., 2008; Adolphs, 2006).

The fixations on specific facial features are used to decode the emotional status of face expression. Among them, the happy expression is the least confused and best-recognized expression when participants locate their fixations on the mouth region (Cangöz et al., 2013; Schurgin et al., 2014). As mentioned before, the happy expression has a special character with a unique upward mouth feature and is also the only positive expression in the basic set of emotions (Leppänen and Hietanen, 2004; Calvo et al., 2010). In contrast, other facial expressions such as anger, fear, sadness, or disgust share one or more similar facial features (see AUs in table 2.1). Fixating on a single facial feature is likely to provide insufficient information to make a correct judgment. Numerous studies have reported confusion between fear and surprise when only the eye region is available for identification (Neta et al., 2017; Palermo and Coltheart, 2004), as well as between anger and disgust when showing only the "nose scrunch" (AU9) (Pochedly et al., 2012; Jack et al., 2009). Although several studies have suggested that participants extract information from the eye region when recognizing angry, fear or sad expressions (Cangöz et al., 2013; Schurgin et al., 2014; Eisenbarth and Alpers, 2011), they compared fixations only located in two areas of interests (AOIs) - the eyes and mouth. When comparing more than these two AOIs, the fixations were located more on the upper nose, the upper lip and the eyes for recognizing fearful faces (Schurgin et al., 2014). These discrepancies arise when treating the nose as one AOI, despite the fact that the majority of the fixation (88.03 %) is not only located on the eyes and mouth but also on the nose region in comparison to all other regions on the face. Also, the strength of a facial expression for a particular emotion can influence the exploration pattern and aids the distinguishing between a subtle or extreme expression. (Vaidya et al., 2014), see Figure 2.5.

Several studies demonstrated that in a direct comparison of the eyes and mouth, the majority of fixations is directed more towards the eye region when recognizing fearful expressions and the mouth for happy expressions (Eisenbarth and Alpers, 2011;



**Figure 2.5:** Face stimuli filtered with high and low spatial frequency bands for optimal information distinguishing facial expression from neutral expression in healthy subjects. In subtle emotional faces (on the upper row), the fixation patterns are similar across emotions with the eye region used as the strongest diagnostic information, followed by the nose and the mouth. In the extreme emotional faces (on the bottom row), the distribution of fixations on diagnostic information areas was varied between emotions. In sum, the diagnostic area on the face is not only limited to the eyes and the mouth region. Reprinted from [Vaidya et al. \(2014\)](#), *Cognition*, with permission.



**Figure 2.6:** Fixation distribution for happy, sad, angry and fearful expressions in the congruent condition in the study of [Bombari et al. \(2013\)](#). The color-code bar indicates the most fixated region in red to the less fixated regions in blue. Reprinted from ([Bombari et al., 2013](#)) with permission.

Scheller et al., 2012). However, when the nose region is taken into consideration, the total fixation is located on the center of the fearful face (Bombari et al., 2013, 2009), see Figure 2.6. Bombari et al. (2013) suggested that fixating on the center of the face (i.e., the nose) reflects an enlargement of the visual attention to gathering information from the whole face rather than extracting information from the nose region only. Some studies argue that the nose cannot be considered as an expressive feature (Buchan et al., 2007). Other studies found that the centralized fixation on the face can be explained as a function of processing configural information Miellet et al. (2011); Hsiao and Cottrell (2008). For example, Westerners tend to extract information from the eye region, whereas Easterners extract information from the center of the face (i.e., the nose) (Miellet et al., 2013). In this case, many studies have neglected the relevance of the nose to the perception of facial expression, particularly in recognizing fearful and angry faces (Schurgin et al., 2014; Bombari et al., 2013, 2009). Therefore, it would be more prudent to include the nose area as a diagnostic feature in the data analysis to provide a complete picture of how human decode facial expressions.

## Cognitive strategies for encoding facial expressions

The eye-tracking technology provides a high temporal precision from 50 Hz up to 2000 Hz sampling rate (Nyström and Holmqvist, 2010; Holmqvist et al., 2011), and in most emotional face processing studies 1000 Hz is used as sampling rate (Schurgin et al., 2014; Scheller et al., 2012; Vaidya et al., 2014). Using eye-tracking allows researchers to examine the temporal precision between fixation onset and offset with a millisecond resolution. Therefore, recent studies not only examine the length of the total fixation duration during emotion recognition but also inspect the single fixation duration and location, such as the first and second fixation (Bombari et al., 2013; Eisenbarth and Alpers, 2011; Schurgin et al., 2014; Vaidya et al., 2014). The investigation of early visual processing (e.g., first and second fixations) of emotional expressions is particularly important to understand where participants locate their visual attention for distinguishing different emotions. This hypothesis was confirmed by other studies (Eisenbarth and Alpers, 2011; Schurgin et al., 2014; Vaidya et al., 2014; Bombari et al.,

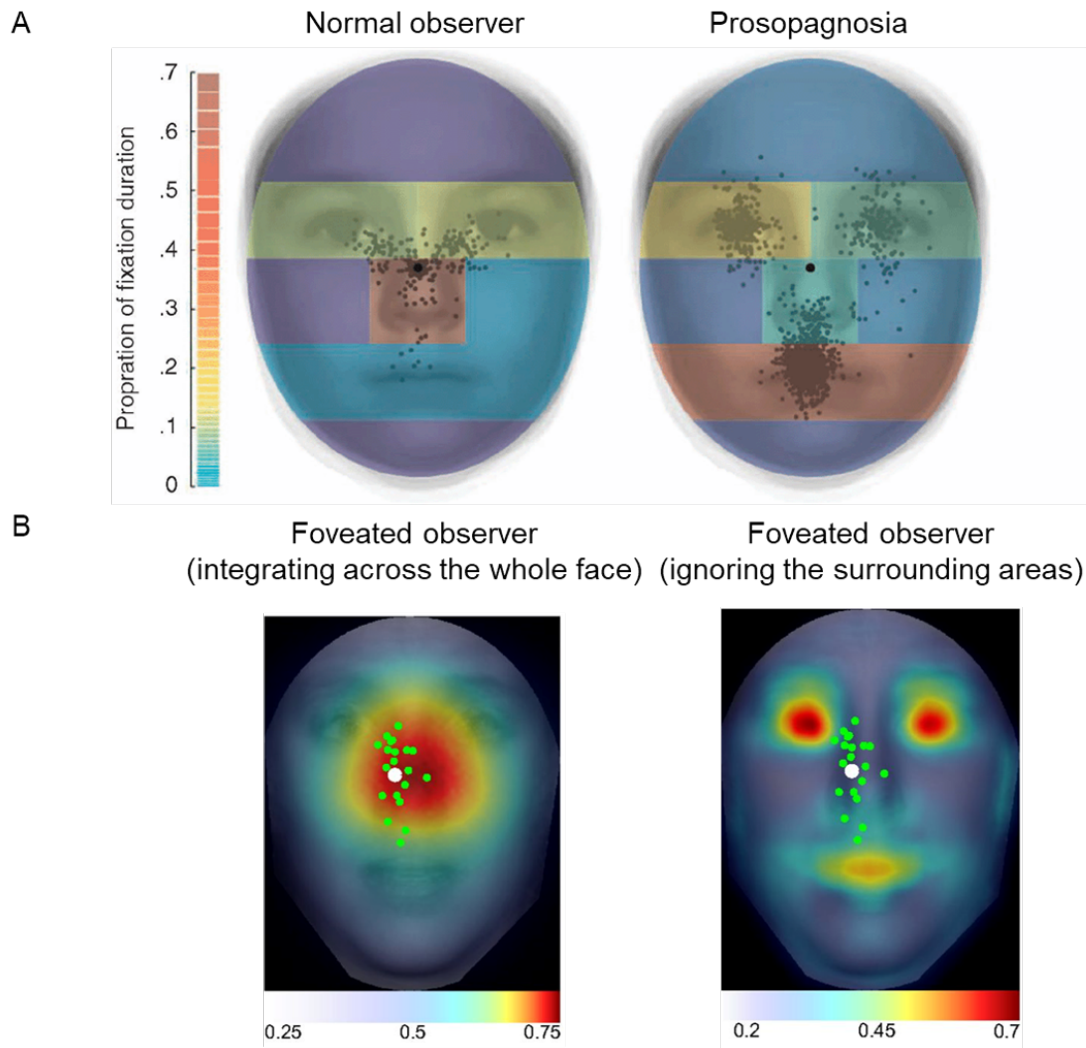
2013).

The first and second fixation and their duration are linked to the function of cognitive processing for decoding facial expressions. [Hsiao and Cottrell \(2008\)](#) has shown that two fixations can be sufficient for recognizing a face. They also revealed that the first and second fixations were always located on the center of the neutral faces regardless the location of the face stimuli or the previous fixation location, suggesting participants were integrating the information into a face configuration. Similar results were also found when participants identified different facial expressions ([Schurgin et al., 2014](#); [Vaidya et al., 2014](#); [Bombari et al., 2013](#)). Some studies suggest that fixating on the center of the face enlarge the parafoveal focus and result in processing information of features without looking directly on them ([Caldara et al., 2010](#); [Richler et al., 2011a](#)). In turn, the distribution of fixations spreads equally from one feature to another feature while the entire face stimuli remain present, which is also associated with configural processing ([Neath-Tavares and Itier, 2016](#)).

The nose may convey subtle expressive information when judge anger and disgust according to the action units ([Ekman and Friesen, 1978](#)), but the role of the nose is rarely discussed in emotion recognition studies. Some studies argue that the nose not carry any expressive information ([Buchan et al., 2007](#)), despite the fact that there is an action unit defined as *Nose wrinkler* (AU9), see Table 2.1. The nose itself might not have much expressive information compared to wide-open eyes or the smiling mouth, but the location of the nose provides the potential possibility of gathering the global configuration. Moreover, if the fixation toward the nose region has no value for decoding facial expressions, then we should expect a low recognition score or no difference across emotions. In order to better understand the role of the nose, in a particular study the participants were asked to identify which nose is belonging to Larry (known as the studies of *Larry's nose*) ([Tanaka and Farah, 1993](#); [Tanaka and Sengco, 1997](#)). They found out that subjects have a 10% increased accuracy in identifying the nose in the context of a whole face (Larry's face with Larry's nose versus Bob's nose) rather than the nose is presented as an isolated feature (Larry's nose versus Bob's nose). In sum, configural processing reflects a potential efficient decoding strategy by fixating on the center of the face.

Fixation patterns can also be influenced by the task itself. For example, [Schyns et al. \(2009\)](#) found that the visual information from the eye region is the determining factor for the identification of a face or its gender. On the other hand, participants extract information from the eyes and the mouth to decide about the emotional valence of a face ([Smith et al., 2005](#)). [Schwarzer et al. \(2005\)](#) studied the gaze behavior in face processing and suggested that participants fixated not only a single facial feature but also located their fixation on the upper part of the face but slightly below the eyes (i.e., the *nasion*), known as the center of the face. In other words, participants looked longer on the nose when the tasks were required configural processing. Some studies concluded that the position on the center of the face might be the place of optimal processing ([Peterson and Eckstein, 2012](#); [Schwarzer et al., 2005](#)). Note that the center of the face refers rather to the mass of the face than to the geometric center because there is more contrast in the upper part of the face ([Rossion, 2014](#)). Supporting evidence come from patients with prosopagnosia who have difficulties with configural perception as they do not fixate on the center of the face (i.e., *nasion* area) as compared to healthy participants ([de Xivry et al., 2008](#); [Peterson and Eckstein, 2012](#); [Rossion, 2014](#)), see Figure 2.7 for further details.

In contrast to the study of [Hsiao and Cottrell \(2008\)](#) examining only the neutral expression, [Bombari et al. \(2013\)](#) found that the first and second fixation duration and locations varied as a function of facial expression. The first fixations were located on the mouth region, and following fixations were oriented to the eye region when recognizing happy expression ([Vaidya et al., 2014](#); [Bombari et al., 2013](#)). However, not all facial expressions share similar visual scanning patterns for the first and second fixations. [Schurgin et al. \(2014\)](#) found that the first fixation and second fixation were located on the nose region when recognizing a fearful expression (see Figure 2.6), whereas [Vaidya et al. \(2014\)](#) found that second fixations were directed more to the eye region for the same expression. Due to the fact that very few studies investigate the influence of configural and part-based processing with fixation temporal dynamic examination, the results are heterogeneous and further complicated by using different face stimuli, number of facial expressions and the task itself. For example, the results of studies of [Vaidya et al. \(2014\)](#) are relatively comparable to the study of [Bombari et al. \(2013\)](#) because both of these studies were using the same face stimuli set (KDEF) examining



**Figure 2.7:** **A.** The distribution of fixations during a personally familiar face recognition task for normal observers (left) and a patient with prosopagnosia (right). The normal observers fixate on the center of the face, slightly below the eyes rather than on any other specific parts of the face. Conversely, a patient with prosopagnosia who has an impaired holistic face perception locates the fixations on each part of the face. 60% of fixations were on the mouth, but also located on each eye (see [de Xivry et al., 2008](#); [Rossion, 2014](#)). Adapted from ([de Xivry et al., 2008](#)) with permission. **B.** The distribution of the mean fixation in an emotion recognition task with overlaid points of mean saccade for each participant (in green) and the group (in white) from two models: 1. The foveated observer (left) who place their fixations on the center of the face, slightly below the eyes, showing where information is optimally integrated from the face. 2. The local observer who extract information from the eye region. (see [Peterson and Eckstein, 2012](#); [Rossion, 2014](#)). Adapted from ([Peterson and Eckstein, 2012](#)) with permission. In sum, the patient with prosopagnosia behave similarly as the local observer as compared to normal observers.

three AOIs (the eyes, nose, mouth) with an emotion labeling task. However, the differences are in the calculation of fixation duration (fixation time compared to the proportion of fixation time) and the number of facial expressions used, see Table 2.2. The error rate increases with the number of facial expression stimuli presented in one experiment (Calvo and Lundqvist, 2008), and potentially modifies the visual exploration strategy for distinguishing between facial expressions (Schurgin et al., 2014). Conclusively, there is a need to examine the first and second fixation processing on emotional faces with a comparable methodology.

**Table 2.2:** Summary of studies analyzing the mean fixation duration, and the first and second fixation duration on healthy subjects.

Study	Task	Emotions and AOIs	Total fixation duration	First fixation duration on an AOI	Second fixation duration on an AOI
Eisenbarth & Alpers, 2011	Valence and arousal rating, n=36. 16 expressors (KDEF) <sup>2</sup> .	Fear, anger, happiness, neutrality and sadness. Eight AOIs: forehead, left eye, right eye, left cheek, nose, right cheek, mouth, chin.	Differed between AOIs and emotions, but the left and right eye region was fixated longer than the forehead, left cheek and nose across emotions.	The location of fixation did not differ from the second fixation.	The left and right eyes were fixated more often for sadness and anger. The mouth region was fixated more for fearful, happy and neutral expressions. The second fixation focused on one of the AOIs more than the first fixation. (Rank: left eye > right eye > mouth > nose > hair > forehead > left cheek > right cheek. No significant difference between eyes and mouth. )
Vaidya et al., 2014	Emotion label and rating extreme level of facial emotion <sup>1</sup> , n=28. 12 expressors (KDEF).	Happiness, fear, disgust, surprise and neutrality. Three AOIs: eyes, nose, mouth	n/a	First fixations (measured as a proportion/frequency of total fixation duration) were located on the mouth for disgust and happiness or surprise. Other features were not targeted.	Second fixations (measured as a proportion/frequency of total fixation duration) were directed to the eye region for fear. No preferred AOI for the other emotions.
Bombard et al., 2013	Emotion recognition <sup>1</sup> , n=26. four expressors, (KDEF).	Anger, happiness, fear, and sadness. Three AOIs: eye, mouth, and nose	The nose region was fixated longer for anger than sadness. The mouth was fixated longer for happiness and anger. The longest fixation duration was found for fearful faces.	The eyes were fixated more for sadness than happiness. The mouth region was fixated more for fear than anger.	The nose was fixated more than the mouth. The eye region was fixated less for happiness than in all other emotions.
Schurgin et al., 2014	Intensity of emotional faces <sup>1</sup> , n=51. 12 expressors in grey scale, (MSFDE) <sup>3</sup> .	Happiness, disgust, fear, anger, sadness, shame and neutrality. Five AOIs: eyes, nasion region, upper nose, lower nose, upper lip.	The eyes were fixated the most across emotions, except for happiness. The nasion region was fixated relatively more for recognising fear and anger.	First fixation was on the upper nose across all emotions, however fixation on the lower nose region varied with emotion.	Second fixation on upper lip for happiness and disgust. Longer fixations on nasion region than on other AOIs.

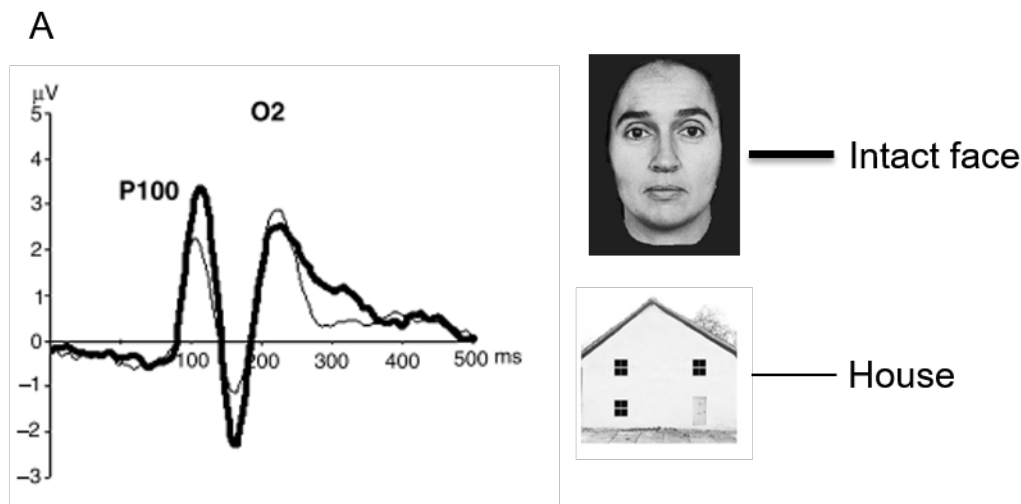
Footnotes: 1) central fixation cross was used before stimulus onset. 2) KDEF = Karolinska Directed Emotional Faces set (Lundqvist, Flykt, & Öhman, 1998). 3) MSFDE = Montreal Set of Facial Displays of Emotion (Beaupré & Hess, 2005).

### 2.4.3 Electroencephalographic (EEG) approach

Electroencephalography (EEG) is a non-invasive measure of the brain activity. Thereby the electrical activity of a human brain will be measured by an electrode on the scalp. The obtained signal will be amplified and the changes in voltage over time evaluated. This remarkable technique was reported by Hans Berger in 1929 ([Berger, 1929](#)) and a few years later confirmed by ([Adrian and Matthews, 1934](#)), [Jasper and Carmichael \(1935\)](#), and [Cole and Ray \(1985\)](#), leading to the acceptance of EEG as a real phenomenon. The EEG measurements allow reflecting the neural responses to specific sensory, cognitive and motor events. These responses can be obtained using a sample averaging technique and reflect electrical potentials that are related to specific events denoted as *event-related potentials* (ERPs) (see [Luck 2012](#)).

The neural electrical activity changes rapidly over time and has a spatially extended field. The advantage of ERPs is their high temporal resolution of a few milliseconds over multiple scalp locations (e.g., electrodes). Moreover, ERP data involves little data processing and temporal filters, each millisecond precision of electrical activity is estimated. The three important aspects of ERP usually monitored are (1) time course, (2) amplitude, and (3) distribution across the scalp. In the context of face perception, researchers found a particular ERP waveform at 170 ms when presenting a face stimulus or showing an object to participants. This evidence shows that the visual information can be followed back to its neural activity (see [Handy 2005](#)). ERPs provide an accurate estimate of the post-stimulus response showing that the brain can distinguish different visual categories. Moreover, ERPs reveal the stimulus-locked response over cortical areas, which are suitable for investigating the visual evoked to the response of face stimuli.

ERPs reveal the timing of neural events underlying in sensory and cognitive processing with millisecond precision across the whole scalp. EEG studies suggest that the perception of visual stimuli (e.g., faces and objects) induces a sequence of evoked components within 200 ms after stimulus presentation ([Rossion, 2014](#); [Bentin et al., 1996](#)). There are two important components in face perception processing, the P100



**Figure 2.8: A.** Grand-averaged ERP waves of P100 components for intact faces and houses on O2 channel. The P100 amplitudes behave differently in the comparison between the face stimuli and house stimuli, the P100 is relatively smaller for houses than for intact faces. The image adapted from (Herrmann et al., 2005), *Journal of Neural Transmission*, with permission.

and the N170 over posterior cortex (see review Luck et al. 2000). The components can be modulated in latency or/and amplitude by different face stimuli such as inverted faces (Sagiv and Bentin, 2001; Itier and Taylor, 2002), sketch faces (Sagiv and Bentin, 2001; Bentin et al., 2006), degraded faces (Linkenkaer-Hansen et al., 1998), and spatial frequency filtered face stimuli (Vuilleumier and Pourtois, 2007; Bortolon et al., 2015; Goffaux et al., 2003).

## The ERP component - P100

P100 (P1) is an early visual ERP component that peaks as a positive component at around 100 ms after stimulus onset and is sensitive to low-visual features of the visual stimuli, such as color, luminosity, contrast or spatial frequency (Liu et al., 2002; Regan, 1989). Also, the P100 component is enhanced by selective attention and sensitive to deviations from a prototypical face (Luck et al., 2000; Halit et al., 2000). Increased P100 amplitudes were seen for example when participants compared faces to houses (Negrini et al., 2017; Herrmann et al., 2005) (see Figure 2.8), for contrast polarity negative faces (Itier and Taylor, 2002, 2004a), for stretched featured faces (Halit et al., 2000), for inverted faces (Boutsen et al., 2006).

There is a debate about if changes in the P100 at the occipital region is elicited by configural processing. While some studies have revealed a shorter P100 latency for upright face stimuli in comparison to inverted face stimuli (known to result in reduced configural processing) (Itier and Taylor, 2002; Linkenkaer-Hansen et al., 1998; Herrmann et al., 2005), other studies describe a similar behavior of the P100 to face or non-face stimuli (Boutsen et al., 2006; Sagiv and Bentin, 2001; Rossion et al., 2003, 1999). Some studies showed that face inversion produced delay P100 latencies and larger P100 amplitudes as compared to upright faces for children older than four years (Taylor et al., 2004) and in adults (Linkenkaer-Hansen et al., 1998; Henderson et al., 2003). Furthermore, several studies have found that the P100 is modulated by spatial-frequency filtered face stimuli (Meaux and Vuilleumier, 2016; Vuilleumier and Pourtois, 2007), and neutral sketch faces (Latinus and Taylor, 2005; Bentin et al., 2006). Critically, the coarse input of face stimuli produced a right lateralization enhancement of the lateral occipital P100 (Pourtois et al., 2005). These results suggest that the P100 may be sensitive to the configuration of a face, presumably for the detection of a face rather than identify the features. In other words, the relevant features must first be located in regions that do not violate the structural description of a face as a typical face. Besides, Herrmann et al. (2005) suggested strongly that the P100 component is associated with an early categorization processing rather than in the actual processing of individual features of faces. However, increased and late P100 components have not been found in sketch faces relative to photographed faces when the task involves emotion recognition.

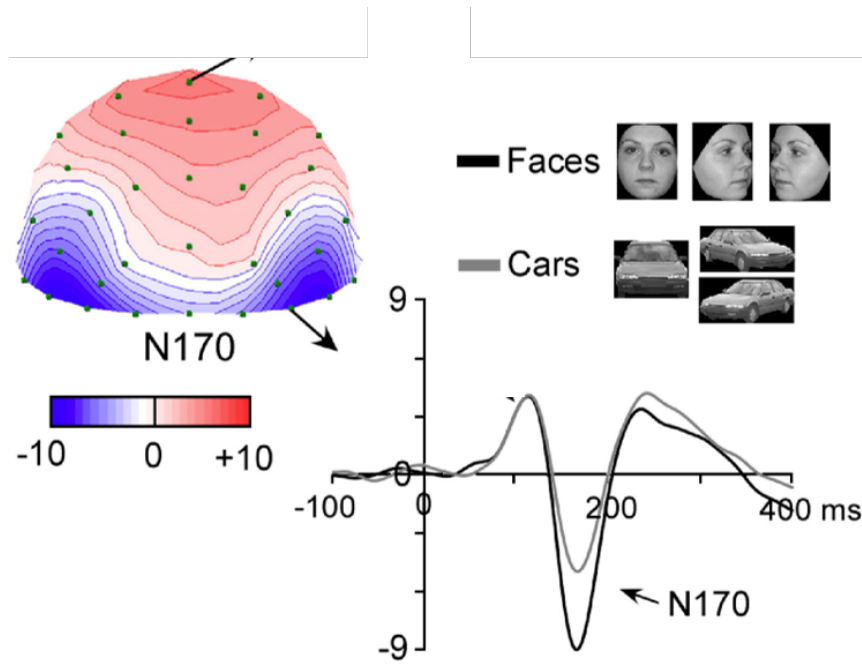
In sum, the P100 component is not only highly sensitive to low-level visual parameters (e.g., luminance, spatial frequency, the color of the face stimulus) but also influenced by the global properties of a face configuration. In turn, the P100 is supposed to be less affected by the content of internal facial features, such as the facial expression. P100 reflects a "face template" matching that eyes above the nose, and the nose above the mouth (Herrmann et al., 2005). Evidence derived from a change in P100 was not observed on both high-spatial frequency and low-spatial frequency faces when facial expressions are engaged (Pourtois et al., 2005). This finding suggests that the early P100 component is not driven by differences in low-level visual features in terms of facial features but rather stimulus's facial texture or global configural information.

There is a demand for a detailed analysis of the P100 component to gain more insights in its role of face processing (Herrmann et al., 2005). Therefore, the P100 component becomes in this work with a cognitive marker to examine if sketch face stimuli are processed similarly to their photographed face counterparts.

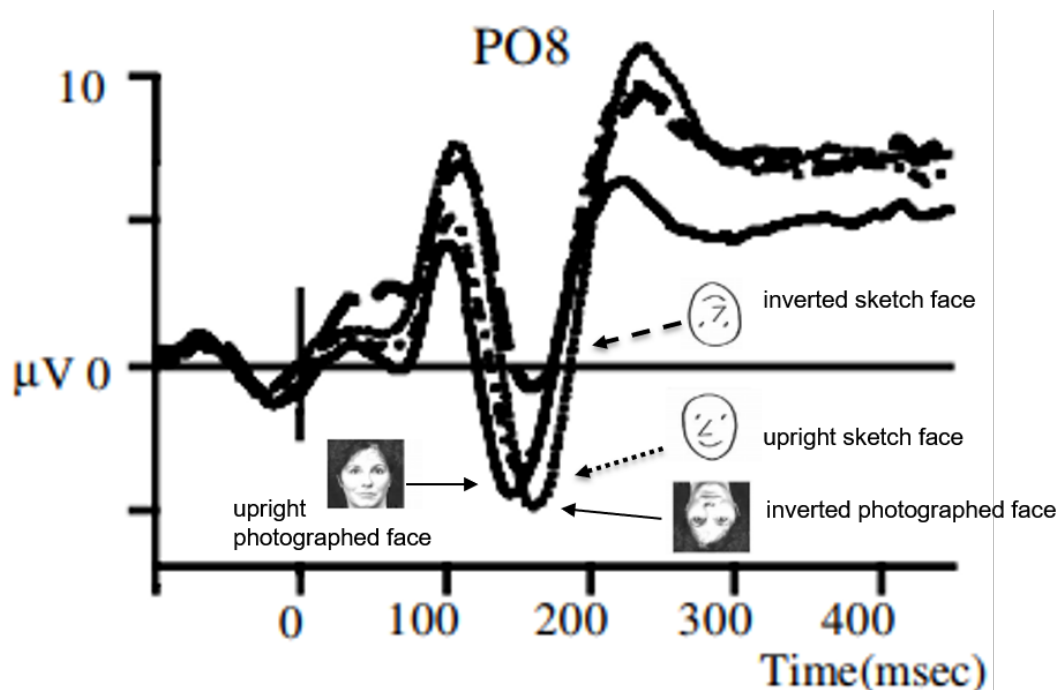
## **The ERP component - N170**

The N170 is a face sensitive component that peak as a negative deflection of the wave amplitude between 120-200 ms after stimulus onset over visual cortex (occipito-temporal sites) (Bentin et al., 1996; Bötzel et al., 1995). The N170 amplitude has shown to be larger in presence of face stimulus than for non-face stimulus (Itier and Taylor, 2004b; Latinus and Taylor, 2005) and discriminates between various face stimuli, such as two tones Mooney faces (incomplete two-tone representation of faces) (Latinus and Taylor, 2005; Gendron and Barrett, 2017), sketch faces (Sagiv and Bentin, 2001; Bentin et al., 2006; Latinus and Taylor, 2006; Kendall et al., 2016), composite faces (Jacques and Rossion, 2009) (see review Rossion 2013), eyeless faces (Itier et al., 2007; Nemrodov et al., 2014; Itier et al., 2011), and spatial frequencies faces (Flevaris et al., 2008; Pourtois et al., 2005; Schyns et al., 2009). However, N170 is not affected by the familiarity of faces (Eimer, 2000a; Bentin and Deouell, 2000), indicating that the N170 component is associated with perceptual face processing that precede in the identification of individual faces. The sensitivity of N170 to faces in comparison to objects like cars is shown in Figure 2.9.

Furthermore, N170 is sensitive to the configuration of a face. For example, the peak latency is delayed and increased for inverted photographed faces (Itier and Taylor, 2002). This ERP inversion effect has been argued to be caused by the disruption of configural processing in inverted faces (Rossion et al., 1999). However, enhanced N170 has not been found in inverted sketch faces, and only a delayed N170 was observed (Henderson et al., 2003; Sagiv and Bentin, 2001), see Figure 2.10. Additionally, the N170 is neither delayed nor enhanced for inverted Mooney faces indicating the absence of configural processing (Latinus and Taylor, 2005; Gendron and Barrett, 2017; George



**Figure 2.9:** The topography (on the left) shows the negative component of N170 activates at posterior lateral electrode sites following faces and car condition (here is car condition). The N170 component peaks at about 160-170 ms following stimulus onset. The N170 is larger in responses to faces than cars in the right hemisphere. Reprinted from [Rossion and Jacques \(2008\)](#), *NeuroImage*, with permission.



**Figure 2.10:** The N170 component elicited at the right hemisphere (08 channel) by upright and inverted photographed faces and sketch faces. A smaller 170 amplitude can be observed for inverted sketch faces than upright sketch and upright photographed and inverted sketch faces. The image adapted from the figure 3 in [Sagiv and Bentin \(2001\)](#), *Journal of cognitive neuroscience*, with permission.

et al., 2005). Yet, the difficulty of recognizing faces is greater for sketch and Mooney faces than for photographed or natural faces (Latinus and Taylor, 2006). Sketch and Mooney faces do not show the N170 ERP inversion effect, which might be due to the ambiguity of the facial features when the stimuli were inverted (Caharel et al., 2013). These results support that the N170 is evoked by configural processing for upright face stimuli, while the inversion disrupts this processing. Latinus and Taylor (2005) argued that Mooney faces rely primarily on the configural information as features are often not distinguishable, whereas inverted sketch faces reflect less association to its configural information than photographed faces (Sagiv and Bentin, 2001). Sagiv and Bentin (2001) suggest that photographed faces carry physiognomy information whereas sketch faces do not, which is one reason for the absence of the face inversion effect on sketch faces. Noted that these findings are established by a face recognition task using the neutral expression not for recognizing a facial expression. Because recognition of specific emotions is not always required to access whole face configural information, such as identifying the happy expression by looking at the mouth region. Along these lines, it is unclear how emotional faces are processed when face configuration is diminished. It is necessary to examine the emotional face processing by measuring the N170 ERP with both inverted and upright emotional face stimuli. Presumably, the features play a critical role in the configural information for the use of face processing.

### **The N170 component as a perceptual expertise biomarker**

Numerous studies have concluded that the N170 reflects the structural decoding of the face. However, some studies suggested that the N170 varies with people and their experiences. This means that a person with cumulative experiences in a specific field is capable of discriminating certain objects better among other categories (e.g., car experts) (Rossion et al., 2003). Carmel and Bentin (2002) suggested that humans process human faces faster than all other visual categories, including ape faces or inverted human faces. Similar to the face inversion effect also a delayed N170 was observed for inverted objects, such as inversion of cars (Rossion et al., 2003), and high-spatial frequency cars (Goffaux et al., 2003). Although some studies reported

no inversion effects for the objects tested, such as cars (Boehm et al., 2011), low-spatial frequency cars (Goffaux et al., 2003), and houses (Boutsen et al., 2006). These contrary results of this delay in early visual processing suggested that the visual system is trained by experience to respond to canonical orientation and configuration (i.e., experts) (Tanaka and Curran, 2001). Further evidence was obtained from fMRI studies, which found a higher fMRI activation in the right fusiform face area (FFA) and the right occipital face area (OFA) for car and bird experts when viewing a visual stimulus, which matched their expertise (Gauthier et al., 2000). Similar findings supported the possibility of an expertise-selective N170 response in EEG (Rossion et al., 2002; Xu, 2005). However, in magnetoencephalography (MEG) studies, some studies failed to observe any expertise effect (termed the N170 component into M170), which might result from occipito-temporal sensors that did not face selective (Xu et al., 2005; Liu et al., 2002). These findings have led to the hypothesis whether face processing is involved in the processing of any class of visual stimuli that share a similar configuration or for which the observers have equipped substantial visual experience as expertise. Presumably, humans are tuned to process photographed natural faces than sketch faces. Hence, we should observe a larger 170 component when processing natural photographed faces than sketch faces.

## **Internal facial features enhance the N170 component**

A typical face configuration activates the neural mechanisms in the early categorization of a face representation. Relative to upright faces, a delayed N170 component was observed for inverted faces as well as for faces without internal features (Eimer, 2000b). The attenuated N170 latency suggested that the absence of internal features is equivalent to the delayed N170 found in inverted faces. Eimer (1998) found a delay in the N170 when using faces without eyes as stimuli. Other studies also found this effect and suggesting that the N170 is unrelated to the presence of eyes but is more likely to the process in the structural decoding of face components (Itier et al., 2007; Nemrodov et al., 2014). These findings do not suggest to neglect the configural information, but the absence of a face-like configuration elicit only a small N170 effect (e.g., the

information of the eyes consequently reflect if the visual stimulus represents a face configuration). A more in-depth investigation of the role which eyes are playing would help to explain why the delayed N170 latency was observed for the face without eyes. It is possible that the presence of eyes within the face accelerating the face processing. Concerning the role of the eyes in speeding up the N170, inversion effects of Mooney and sketch faces are in agreement with a decreased N170 amplitude due to the vague information of the eye features (Latinus and Taylor, 2005; Sagiv and Bentin, 2001; Henderson et al., 2003). It corresponds to the vital role of eyes in emotion recognition that provides direction as a cue of active communication and as a fundamental aspect of human social cognition as discussed in the previous section (see chapter 2.4.2).

## **The influence of configural information**

Face detection is believed to be biased by facial features, resulting in an N170 effect elicited even by sketch faces, because the typical face configuration is retained (Bentin et al., 2006). However, it is surprising that the inversion effect on the N170 was not observed in sketch faces. Because it is commonly assumed that the face inversion disrupts the configural processing and set off featural part-based processing (e.g., (Bartlett and Searcy, 1993; Peters and Kemner, 2017; Carey et al., 1977; Flevaris et al., 2008). Interestingly, inversion of different contrast human faces (e.g., negative vs. positive polarity) revealed parallel additive effects of N170, in which a larger N170 amplitude was found for inverted negative faces than for inverted positive faces (Itier et al., 2006). Similar inversion effect was also observed for the sketch and photographed faces (Sagiv and Bentin, 2001; Liu-Shuang et al., 2015). These results demonstrated that face inversion affects more for atypical face stimuli than natural photographed faces that preserve the prototypical spatial layout of the features. Therefore, it explains the N170 is even more delayed when the face is both inverted and a not a natural photographs. Moreover, when a face retains only internal facial feature without face contour generate a larger N170 than a full face, suggesting the featural information is enhanced by removing the face contour (Flevaris et al., 2008; Bentin et al., 2006). The larger N170 reflects the not only the disruption of configural processing but also the

processing of internal facial components. Particular. It is possible that the intermediate complexity of visual features maximizing the delivered information.

## 2.5 Hypotheses

The aim of this thesis is twofold, first to identify which diagnostic features are necessary for recognizing emotions successfully, and secondly when the processing of diagnostic features is more dominant over configural processing. The original idea behind these aims is to acquire the behavior, eye-movement and EEG markers from healthy participants and to investigate how facial expressions are decoded and when does configural and part-based processing take place during emotion recognition. The underlying reason for these investigations is to establish responses and markers from healthy subjects and use them in future experiments with subjects suffering from mental problems. For example individuals with schizophrenia exhibit an aberrant eye-movement causing the fixation on less informative areas and leading to deficits in configural processing ([Morris et al., 2009](#); [Lee et al., 2010](#); [Baudouin et al., 2008](#)).

The here obtained results will foster the understanding of how the neural network will react when engaging in social cognitive processing. Therefore, we created a new set of sketch faces containing five different emotions (angry, sad, happy, fear and neutral), which were transformed from the validated Radbound Faces database ([Langner et al., 2010](#)). We examine if the degraded configural information with isolated facial features on the newly created emotional sketch faces will be processed differently in comparison to their original photographed faces using behavioral measures, Eye-tracking and EEG measurements.

## **Experiment 1: Stimuli design and validation**

In order to examine the role of diagnostic features and configural information in emotion recognition, we first created a set of sketch faces. These new sketch faces containing five different emotions (happy, fear, angry, sad and neutral) and each diagnostic facial features (eyes, eye-brows, nose, and mouth) is isolated on each face and is placed into a face configuration as its photograph counterpart. To our knowledge, this is no validated face databased creating for sketch faces, which is designed for the emotion recognition. Therefore it is important to examine whether participants can recognize sketch faces as faces as well as identify correctly the expressed emotions. Unlike conventional unnatural face stimuli, sketch faces capture the relevant expressive information which is embedded in the facial features, such as the detail on feature's contour and edges without subjectively modify the information (e.g., bias from drawing preference). In Experiment 1 we hypothesized that sketch faces carry sufficient information for facial recognition based on the presence of informative diagnostic features. Still, we had to assume that not all sketch faces can be identified successfully because facial emotions are sometimes ambiguous and we used five different facial expressions in the experiment. The complete sketch stimuli set was validated via behavioral measures and evaluated according to the reaction time and accuracy. Based on these results only the ten best stimuli per emotion (angry, fearful, sad, neutral and happy) were used for the comparison of sketch and photographed faces in the next experiments using eye-tracking and EEG studies.

## **Experiment 2: Fixation patterns of emotional sketch and photographed faces**

In Experiment 2 we investigated the selected sketch faces of Experiment 1 and their photographed counterparts using eye-tracking measurements. If sketch faces carry sufficient information about the facial expression, the outcome of behavior performance and general fixation scanning patterns should be equivalent to the photographed faces

based on their common diagnostic features. However, due to the degradation of the configural information on sketch faces, we expected changes in the temporal organization of fixations toward the facial feature (e.g., AOI), as compared to photographed faces. Meanwhile, this would provide a way to investigate which part of the face is informative when the configural information is less available. To the best of our knowledge, this approach is very original given that only a few studies are examining the role of sketch and photographed face stimuli concerning the emotion recognition with the eye-tracking measures in this extent.

### **Experiment 3: ERP study of inverted sketch and photographed faces**

We theorized that subjects extract expressive information from diagnostic facial features as well as from configural information in order to grasp global information of arrangement between facial features when judging emotions. Based on the literature we further assumed that the disruption of configural processing by viewing inverted faces leads to a lower accuracy and longer response time than for upright faces. Behaviourally speaking, sketch faces should generate a similar performance as photographed faces when presenting in the upright orientation, considering the presence of diagnostic features and previous studies. On the other hand, an inversion might have a more significant impact on sketch faces than on photographed faces considering the configural information are more degraded on sketch faces. Regarding the effect on ERPs, we assume that the characteristic P100 and N170 will behave differently between sketch and photographed faces. Firstly, the P100 should be larger for processing photographed faces than for sketch faces, because the P100 is more sensitive towards configural information. Secondly, the face inversion effect should be impaired on sketch faces because the configural information is reduced on sketch faces. Therefore, we should observe a smaller and more delayed N170 ERP component when processing inverted sketch faces in comparison to the inverted photographed faces. In sum, we assume that the face inversion effect will not be observed for sketch faces as compared to photographed faces, but participants can recognize emotions on both stimuli due to the presence of diagnostic facial features.

## **Experiment 4: ERP study of configural effects on sketch faces**

To get more insights on the influence of the limited configural information on sketch faces, we modulated the sketch face configuration with different degree of deconfigural parameters (i.e., increase distances between facial features linearly). We hypothesized in Experiment 4 that sketch faces depend less on configural information, we might not observe any modulation of the ERP components (P100 and N170). As a consequence, the diagnostic information plays a crucial role (e.g., part-based processing) in emotion recognition. The behavioral performance might not be affected significantly by the deconfiguration.



# Chapter 3

## Experimental results

## 3.1 Validation of sketch faces

The purpose of this experiment was to validate if the newly created set of sketch faces, transformed from photographed faces from the Radboud Faces Database, still convey sufficient information for recognizing the presented emotions (fearful, angry, happy, sad, and neutral expressions). This chapter provides in detail the design of the sketch stimuli and their validation using behavioral measures. The presented data and discussion can be found in a brief version of the supporting information of Publication 1 (see Chapter 3.3).

### 3.1.1 Hypotheses

Assuming that in the created sketch faces the diagnostic facial features are preserved, we expected that participants are able to classify the facial expressions on sketch faces.

### 3.1.2 Methods

**Participants** Twenty (13 male, seven female) participants with normal or corrected-to-normal vision gave their informed consent to participate in this experiment. Their mean age was 28.4 years ( $SD = 6.7$ , age range 19-42). This experiment was conducted in accordance with the Declaration of Helsinki.

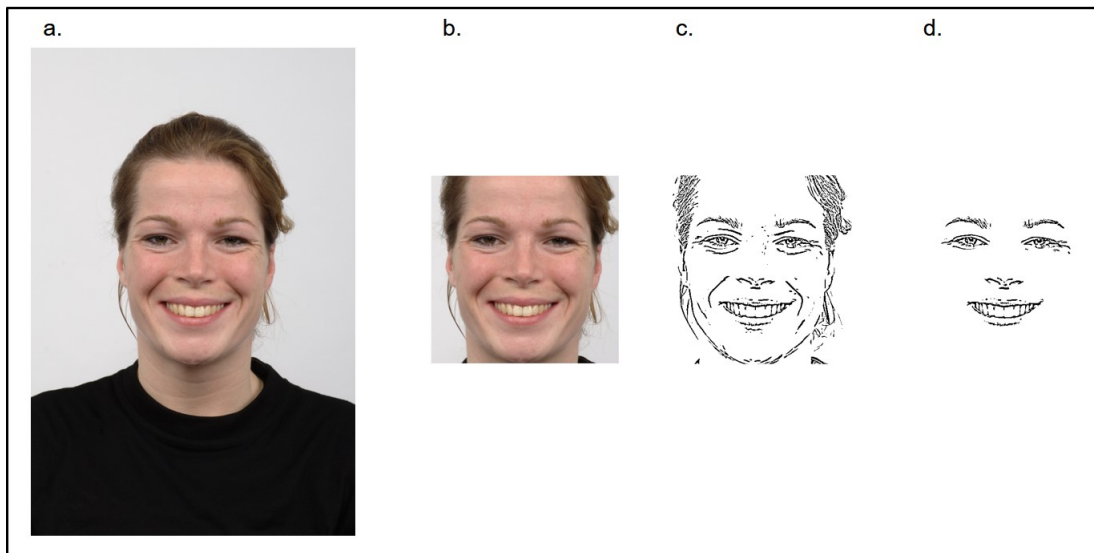
#### Material

**Design of the sketch stimuli** To test whether the internal facial features (i.e., eyes, nose, and mouth) convey sufficient information to recognize universal emotions, we

created a new set of sketch face stimuli by transforming photographed faces from the Radboud Faces Database (RaFD) (Langner et al., 2010). We selected photographed faces of 38 Caucasian adults (19 females, 19 males) from the RaFD. In these, both the head and gaze faced the camera and five emotions were displayed (i.e., fear, anger, happiness, neutrality, and sadness). In the RaFD, eight emotional expressions (including “neutrality”) were validated. However, in our study, we chose four emotions and the neutral expression to limit task length and difficulty. Further, we wanted to avoid using facial expressions, which are known to be easily misinterpreted, such as disgust and surprise (Langner et al., 2010; Goeleven et al., 2008). Therefore, this thesis focuses on the neutral expression and the basic emotions of fear, anger, happy, and sad.

The creation of the emotional sketch face stimuli involved two steps: a preprocessing and a transformation step. In the preprocessing step, a “high-spatial sketch filter” was implemented with OpenCV Library to transform the selected photographed faces of the RaFD database into grayscale sketched facial stimuli by removing facial texture and color. Following, the central facial region was cropped from the original size in RaFD (681 x 1024 pixel) to the size of 400 x 400 pixel. This area roughly corresponded to the facial contours. Figure 3.1 shows a sample of the original photographic happy face stimulus, the cropped photograph and the final sketch face stimulus. The cropped sketch faces were then further transformed to enhance the facial features.

After the preprocessing step the resultant sketch faces were manipulated in a transformation step in which the diagnostic features are extracted. In the extraction step, we aimed to maximize the area of information from the diagnostic facial features while minimizing the differences caused by non-expressive cues (e.g., black spot, shadow). But, a certainly acceptable similarity is necessary ensuring valid measurements between different faces. For this purpose, I delineated the contour of each diagnostic facial feature on each sketch face stimulus and cropped the area based on its face regions manually with an image manipulation program of GIMP2 ([www.gimp.org](http://www.gimp.org)) (see Figure 3.2(a)). Each facial feature is defined in a specific position in the sketch face coordinate, which is also the case for their photographic face counterpart stimuli. Hence, each face stimulus is rigidly defined with the left eye, left eyebrow, right eye,

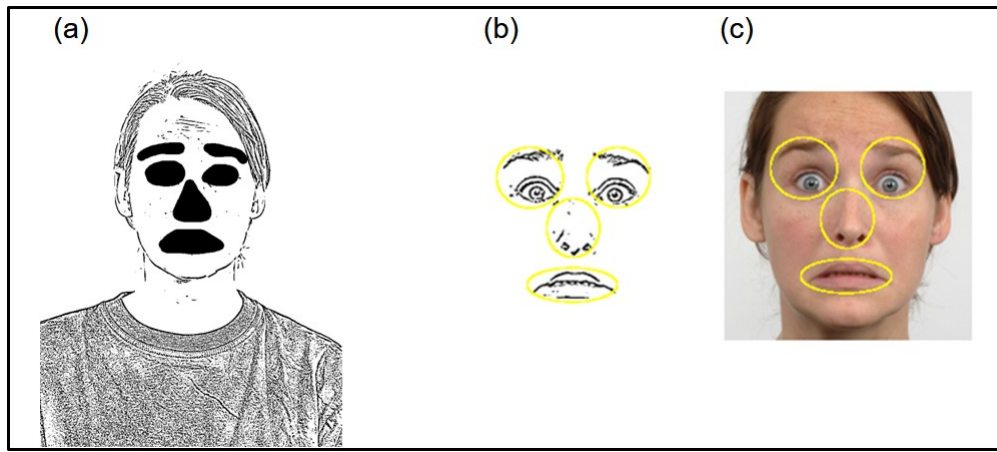


**Figure 3.1:** From left to right: (a) a photograph of a happy female face from the Radboud Faces Database; (b) the central facial region was cropped into 400 x 400 pixel; (c) a sketch-filtered picture; (d) a sketch face with the selection of its diagnostic features (eyes, nose, and mouth), termed as “sketch faces” in this thesis. Note that the size of the images is not the same as used in the experiment.

right eyebrow, nose and mouth, which applies for both photo and sketch face stimuli (see Figure 3.2 (b) & (c)). Each face was then defined as a new face stimulus with those features, for sketch faces the hair, contour, and non-significant details were exempted.

In order to better understand the importance of each diagnostic facial feature and for a better comparison with data from the literature, we divided the facial features into three main groups as our areas of interest (AOIs) - eyes (left eye + eyebrow & right eye + eyebrow), nose, and mouth. Four elliptically shaped regions were defined using EyeLink Dataviewer software with coordinate files (SR Research Ltd) (see Figure 3.2 (b) & (c) for output results). The purpose is not only to create a very precise area of interest per each image but also to remove spurious pixels, which are insignificant to facial features. In the rest of the thesis, these emotional sketch face stimuli are termed as sketch faces.

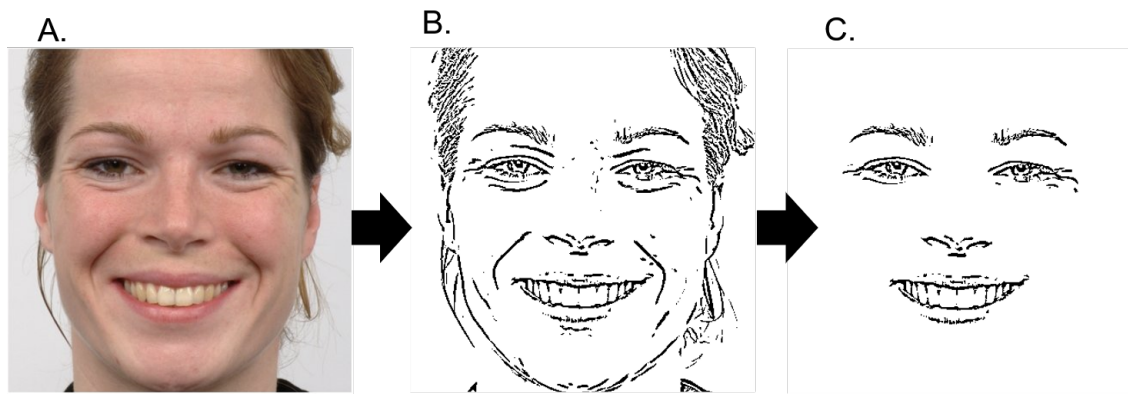
Note that, sketch face stimuli are based on an image filter that reflects the details on the contours of each feature from photograph pictures, as well as discard the noise background. Typical line-drawing or sketch face stimuli used in the literature are drawn by a person. However, this can lead to sketch stimuli which are influenced by subjective bias. The advantage of the here used computer-based transformation is



**Figure 3.2:** Example of pre-defined areas of interest (AOIs) for a fearful female expressor: (a) several areas were defined manually on the RaFD stimulus: left and right eye, both with eyebrows, nose, and mouth, (b) Outputs of sketched, and (c) photographed faces with elliptical AOI shapes defined using EyeLink Dataviewer software (SR Research Ltd).

the creation of a sketch stimulus which represents the photographic face counterpart without personal bias. This strategy is uniform and could also be used for other stimuli. It could be interesting for studies, which want to compare the effect of texture and shape information on the perception strategy. Faces, objects, and situations can be presented in a rich version and in a hand-drawing-like version in which the amount of texture information can be fine-tuned by the settings in the preprocessing and transformation step. Making this approach to a versatile objective method to create hand-drawing-like stimuli.

**Experimental Stimuli** We selected frontal view photographs of 38 Caucasian adults (19 female, 19 male) from the Radboud Faces Database ([Langner et al., 2010](#)). Both the head and the gaze faced the camera and five emotions were displayed (i.e., fear, anger, happiness, neutrality, and sadness). Creation of emotional sketch face stimuli involved two steps: pre-processing and transformation. In the pre-processing step, a “sketch filter” was used with OpenCV Library to transform stimuli in the photographs into grayscale sketch facial stimuli. Then, the central facial region was cropped from the original RaFD 681 × 1024-pixel area to a 400 × 400-pixel area corresponding roughly to the facial contours (see Figure 3.3 A and 3.3 B). The purpose of the transformation step was to keep areas corresponding to the diagnostic facial features, so-called areas of interest (AOIs): eyes (left eye & eyebrow + right eye & eyebrow), nose, and mouth



**Figure 3.3:** From left to right: (A). a cropped 400 × 400-pixel photograph of a happy female face from the Radboud Faces Database, (B). a sketch-filtered picture, (C). sketch face with diagnostic information for the eyes, nose, and mouth (e.g., AOI)

(see Figure 3.3 C). AOIs were delineated manually for each stimulus. Elements such as wrinkles or black mole/spot were excluded for the AOIs. Figure 3.3 illustrates the steps leading to the sketch face stimuli.

## Apparatus

The participants were tested individually in a quiet room. Stimuli were presented on a 17-inch monitor (resolution 1280 × 1024 pixel, screen width is 340 mm × height is 270 mm) with a 60 Hz refresh rate. The size of the stimulus images was 400 × 400 pixels, which equals to 10° × 10° visual angle at a view distance of 60 cm. The value of the visual angle resembles a normal face-to-face interaction (Henderson et al., 2005; Hsiao and Cottrell, 2008). E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used to present stimuli and record behavioral performance.

## Procedure

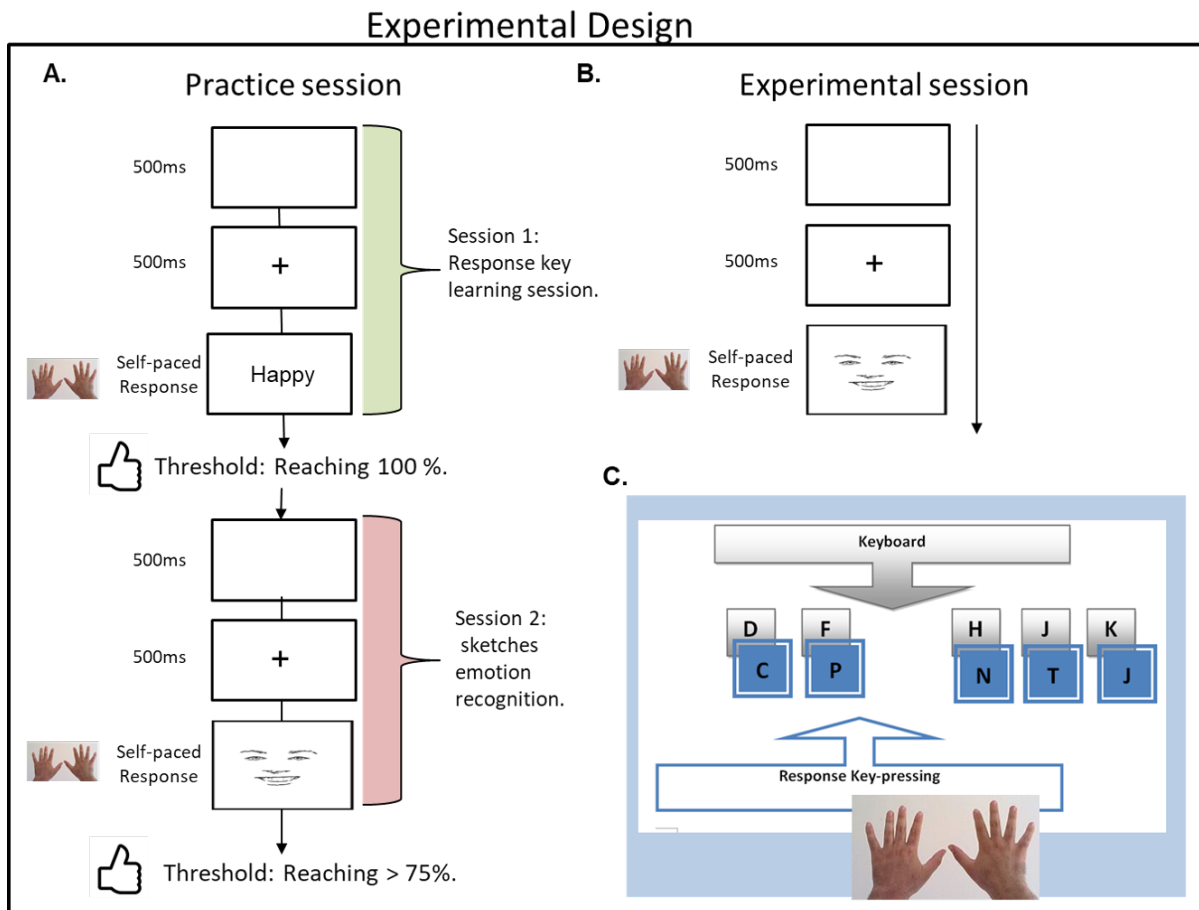
The experiment involved a practice session and a experimental session, see Figure 3.4. The practice session was subdivided into session 1: a response key learning session and session 2: a sketch emotion recognition session (Figure 3.4 A). Each practice

session involved a response feedback about the performance. In each session (practice or experimental), the trial began with a 500 ms fixation cross, after a 500 ms blank screen, followed by the stimulus, which remained on the screen until the participant responded (self-paced). The task was self-paced with no time limit. The participants were instructed to respond as quickly and accurately as possible.

Response key learning Practice session 1: This session ensured that the participants were able to correctly associate a given response key with a specific emotional state. The “D”, “F”, “H”, “J” and “K” keys of an AZERTY keyboard were used, with these keys re-labelled with stickers as follows: “C” (for “Colère”, meaning “anger” in French); “P” (for “Peur” = fear); “N” (for “Neutre” = neutral), “T” (for “Tristesse” = sadness), and “J” (for “Joie” = happiness), see Figure 3.4 C. This session consisted of a minimum of 30 trials where in each trial a randomly selected emotional state was displayed as a word on the screen to which participants responded by pressing the corresponding key. After each five-trial block, the average response accuracy (in %) was displayed. This response key learning session continued until the participants attained 100% accuracy over five successive blocks.

Sketch emotion recognition practice session 2: each trial consisted of one stimulus (either a female or male expressor) presented in the centre of the screen with a random order of the five facial expressions. The participants pressed the corresponding key in response to the expression being displayed, see Figure 3.4 A. After each block of five trials, the average response accuracy (in %) was displayed. There were six blocks in total. Participants obtaining less than 75% accuracy on average were excluded from the rest of the study. The practice session contained ten images (two face identities with five facial expressions), that were not used in the experimental session.

In order to continue with the experimental session, the participants were required to reach 100% accuracy in session 1, then 75% accuracy in session 2. The procedure for the experimental session was the same as in the practice session 2 except that the sketch stimuli were different (36 expressors per each emotion, 180 stimuli. 2 expressors used in the practice session were excluded from experimental session),



**Figure 3.4:** This scheme illustrates the experimental design. (A). Practice session subdivided into two sessions, key learning practise session (top) and emotion recognition practice session using the actual face stimuli (bottom). (B). The paradigm design of the experimental session. (C). Relabelled French AZERTY keyboard (in upper row) with each initial letter of targeted expression (in bottom row). Participants were encouraged to use two hands to choose responded key.

and the participants received no feedback on response accuracy, see Figure 3.4 B. Each session included 180 randomized sketch faces (36 face identities  $\times$  five facial expressions). There were four blocks of 45 trials separated by self-paced pauses.

## Data Analyses

Unbiased hit rates ( $Hu$ ) and reaction times (RT) for correct response trials were analyzed for each stimulus. Thereby,  $Hu$  were computed for each emotional state category. Wagner (1993) argued that using the unbiased hit rate is the most conservative measure of accuracy in a categorical judgment study (see studies that have also used  $Hu$ :

Beaupré and Hess, 2005; Biehl et al., 1997; Elfenbein and Ambady, 2003; Goeleven et al., 2008; Langner et al., 2010; Leppänen and Hietanen, 2004). In general, the  $Hu$  vary between 0 and 1 and reveals whether an accuracy increase was due to bias (i.e., chance level of guessing) or due to an increased sensitivity of target emotion labelling. The calculation of  $Hu$  takes into account both the hit rate and false alarm rate (see Equation 3.1). The former describes the proportion of correctly categorized target emotion trials whereas the latter is the number of missclassified target emotion trials.

$$Hu = Hits^2 / (AiBi) \quad (3.1)$$

In the Equation 3.1, Hits is the frequency of corrected answer trials,  $Ai$  the number of trials (hits and false alarms) per response condition where  $i$  is the target emotion and  $Bi$  is the number of trials (hits and false alarms) per stimulus condition where  $i$  is target emotion in this case.  $Hu$  is then normalized using an arcsine transformation, leading to a variation of the value between 0 and 1. Furthermore, Bonferroni-corrected t-tests were conducted to test our conservative criterion ( $Hu = .75$ ) and chance level of response ( $Hu = .20$ ) for each emotion.

Trials with a RT greater than “mean + 3 SD” (9515 ms) were excluded from the analysis, which corresponds to 1.8% over all trials. Furthermore, trials with a RT between “mean + 2 SD” and “mean + 3 SD” were replaced by “mean + 2 SD” (7039 ms), corresponding to 1.63% of all trials. Repeated-measures ANOVAs were conducted on RT and  $Hu$  with emotion as a within-subject factor (fear, anger, happiness, neutrality, and sadness), followed by post-hoc Tukey (HSD) comparisons when the main effect was significant. Greenhouse-Geisser corrections were applied when the sphericity assumption was violated ( $\epsilon$  values are given in this case). Moreover, t-tests were performed using the Bonferroni correction method to determine whether  $Hu$  exceeded the chance level of .20, and the criterion for the accurate recognition of  $Hu = .75$ . A detection performance level of .75 is rather conservative given that it is classically used to define a threshold stimulus level for satisfactory detection in two-alternative forced choice psychophysical paradigms (i.e., with a chance level of .50 for recognition).

However, the chance level was in our studies initially much lower.

### 3.1.3 Results

The behavioural data collected during this first experiment with the new created sketch stimuli and their photographed face counterparts was analyzed in two steps. A first analysis was performed considering all 180 stimuli (36 Face Identities x five Facial Expressions), termed sketch faces database. Based on the results of the first analysis a subset of 50 stimuli (10 Face Identities x five Facial Expressions) were selected, which showed a high accuracy in their recognition. This subset was further analyzed in a second analysis step.

#### Behavioural performance for the sketch face database (180 stimuli)

The unbiased hit rate varied significantly with emotion,  $F(4, 76) = 136.53, p < .0001$ ,  $\eta^2_p = .88$ . Post-hoc tests indicated that only anger and sadness did not differ for  $Hu$  ( $p = .99$ ); all other pairwise comparisons were significant ( $ps < .008$ ) (see Table 3.1). The mean  $Hu$  for each emotion was significantly above the chance level (0.20); all measured  $ps < .01$  (using  $t(19)$  tests with a Bonferroni correction of  $\alpha = 0.05/5$ ). A satisfactory recognition rate should be greater than  $Hu = 0.75$ ; thus, we ran additional Bonferroni-corrected comparisons that showed that the recognition of anger, neutrality, and sadness was significantly below this threshold (all  $ps < .01$ ), whereas recognition of fear and happiness was significantly above it (both  $ps < .01$ ). RTs also varied significantly with emotion ( $F(4, 76) = 26.52, \epsilon = 0.56, p < .0001, \eta^2_p = .58$ ). Post-hoc tests showed that the participants recognized angry faces more slowly ( $ps < .001$ ), and happy faces more quickly ( $ps < .0001$ ) than the other emotions. Neutral expressions and those of sadness and fear did not differ significantly with regard to RTs ( $ps > .93$ ) (see Table 3.1).

**Table 3.1:** Summary of RT and *Hu* per emotion for sketch faces (n=180) database.

Measure	RT (ms)			<i>Hu</i>		
	<i>M</i>	95% CI		<i>M</i>	95% CI	
		<i>LL</i>	<i>UL</i>		<i>LL</i>	<i>UL</i>
Angry	2268	2126	2410	.49	.42	.55
Fearful	1768	1620	1976	.85	.79	.91
Neutral	1862	1749	1901	.59	.55	.62
Sad	1823	1745	1901	.48	.45	.55
Happy	1052	1018	1087	.96	.94	.97

*Note.* *M* = mean, CI = confidence interval; *LL* = lower limit, *UL* = upper limit.

### Behavioural performance for the sketch face subset (50 stimuli)

Although the *Hu* of the sketch faces database was above the chance level, we assumed that well-recognized sketch face stimuli should lead to a *Hu* greater than .75 to carry emotion-specific content. Accordingly, we ranked the *Hu* and RT performance for each stimulus to select ten face identities per emotion as a subset of the sketch face stimuli database. The *Hu* was used as first criterion and RT was used as second criterion. In order to balance the gender of face identity per emotion, the top five face identities per gender per emotion were selected. Bonferroni-corrected comparisons showed that the recognition of anger, neutrality, fear, happiness and sadness was significantly above the  $Hu > .75$  threshold (all  $ps < .01$ ). As expected, *Hu* for the sketch faces subset was greater than the sketch faces database ( $ps < .01$ ) (see Table 3.2).

The repeated-measures ANOVA on the *Hu* revealed a significant effect depending on emotion ( $F(4, 76) = 11.87$ ,  $\epsilon = 0.61$ ,  $p < .0001$ ,  $\eta^2_p = .38$ ). Post-hoc comparisons showed that anger and sadness did not differ with regard to *Hu* ( $p < .5$ ), see Table 3.2, and were similar to the sketch faces database result. However, the levels of accuracy for the fearful and the neutral expressions were closer to that of happiness (i.e., not significantly different, with  $p = .45$  and  $p = .49$ , respectively). A repeated-measures ANOVA on the RTs also showed a significant main effect of the emotions ( $F(4, 76) = 13.41$ ,  $\epsilon = 0.62$ ,  $p < .0001$ ,  $\eta^2_p = .41$ ). Further analyses demonstrated that there was still an advantage for happy faces (recognized more quickly than the other emotions, all  $ps < .0001$ ). In comparison to the complete sketch database the recognition of angry

**Table 3.2:** Summary of RT and *Hu* per emotion for the sketch faces (n=50) subset.

Measure	RT (ms)			<i>Hu</i>		
	<i>M</i>	95% CI		<i>M</i>	95% CI	
		<i>LL</i>	<i>UL</i>		<i>LL</i>	<i>UL</i>
Angry	2205	1727	2682	.87	.81	.92
Fearful	1625	1283	1966	.96	.93	.99
Neutral	1831	1537	2126	.96	.92	.99
Sad	1773	1529	2017	.90	.86	.94
Happy	931	850	1002	.99	.98	1

*Note.* *M* = mean, CI = confidence interval; *LL* = lower limit, *UL* = upper limit.

faces improved so that RTs were now similar to those for neutral and sad expressions ( $ps > .12$ ) but longer than for fear ( $p = .0001$ ). RTs still showed no significant difference between fear, neutral expression, and sadness ( $ps > .78$ ).

### 3.1.4 Discussion

This study investigated whether sketched pictures of faces can convey sufficient social information to enable the recognition of the presented four basic emotions and neutral facial expressions. The sketch face database set (180 stimuli) showed significant differences in terms of emotion recognition. High recognition performance was observed for happy and fearful expressions. Although, the accuracy of angry, sad, and neutral expressions were above the chance level (0.20), but not above the satisfactory recognition rate ( $> .75$ ). Thus, we selected a sketch face subset (50 stimuli) for further investigation. This subset was selected according to the value of the unbiased hit rate *Hu* (needed to be above a strict threshold of .75) and short reaction time. The subset was then analyzed again and could confirm the better performance on average in comparison to the complete sketch dataset. Although these stimuli are somehow impoverished with respect to the originally photographed faces, we assumed that sketch faces would increase the salience of diagnostic facial features, for instance, facial edges or contours and thus may be more easily processed by the visual system. Happy expressions were recognized faster and more accurately than other emotions (sadness, fear, anger, and neutrality). This happy face advantage is consistent with earlier studies

that have used photographs or line drawing stimuli ([Eisenbarth and Alpers, 2011](#); [Horstmann and Bauland, 2006](#); [Leppänen and Hietanen, 2004](#)). We further noticed that a fearful face was recognized more accurately than an angry expression.

Sketch faces capture the most informative part of the face in a concise and robust face representation. However, generating a clear facial expression can be challenging due to the lack of rich facial details and facial structural variations. In this study, sad and angry expressions were more difficult to identify. Similar results were found in a different study which was using peripheral recognition and could show that angry, fearful, and sad faces need to fall within the central vision (i.e., foveated) in order to reach a better performance ([Goren and Wilson, 2006](#)). The processing of face stimuli in the periphery is believed to reduce configural processing. This evidence suggests that sketch faces degrade the configural information, resulting in increasing difficulty in identifying anger and sad faces, but aiding the recognition when there is a highly distinctive feature (happy expressions).

However, the role of the fearful face is unclear here; some studies suggest that configural information are crucial for recognizing a fearful expression [Bombari et al. \(2013\)](#); [Prkachin \(2003\)](#). Although we found a good performance for the sketch face database set, the high-edge contour of sketch faces might facilitate fearful face recognition. Typically the wide-open eye expression is the most prominent diagnostic facial feature of a fearful face. [Whalen \(2004\)](#) found that the human amygdala responded more strongly to isolated black-and-white fearful eyes similar to those shown in our sketch faces. In how far facial components are modulating the recognition performance of emotional facial expressions within sketch faces is the topic of the next experiment. Overall, the results showed that emotions could be recognized from sketch facial stimuli, albeit they contain less rich visual information than the photographed faces counterparts of the RaFD. As the performance varied by emotion, it will be interesting to use eye-tracking to investigate whether or not the visual exploration strategy of the sketch face stimuli differs from their photographic counterparts.

## 3.2 Fixation patterns of sketch and photographed faces

After the validation of the new created sketch face stimuli, we used the subset of selected sketch face stimuli with the highest performance in emotion recognition in a comparison study with their photographed face counterparts. For the comparison we were interested in similarities and differences between both stimuli regarding the fixation location and duration measured with eye-tracking. As this experiment is the major subject of the Publication 1, which is provided in the full length with supplementary information in the following chapter (see Chapter 3.3), this chapter is intended to give only a brief summary about the most important details of the hypothesis and discussion.

### 3.2.1 Hypotheses

The results of the previous experiment, described in Chapter 3.1, have shown that the selected sketch stimuli could convey sufficient expressive information for successful recognition of the emotional status. Based on previous investigations in the literature two main visual processes were identified for the recognition of facial expression, namely part-based and configural processing (Meaux and Vuilleumier, 2016; Bombari et al., 2013; Wegrzyn et al., 2017). Sketch faces were suggested to trigger more part-based processing than photographed faces. The reduced configural information might reveal the role of diagnostic features on sketch faces. Therefore, in this experiment, we were interested to compare the visual processing for decoding a facial expression between sketch faces and their photographed face counterparts. We compared sketch faces that emphasize the facial features with photographed faces that contain rich visual information (color, shadow, face contour, etc.) by the mean of measuring the location and duration of eye fixations using eye-tracking.

### 3.2.2 Discussion

This comparison study of the two stimuli, e.g., sketch and photographed faces, revealed that both enabled emotion recognition with high accuracy. Thereby we could confirm the results of the validation study in Chapter 3.1 as a high performance was observed for the selected subset of the created sketch faces database, which is essential for the reproducibility of the results.

The results of the behavioral measurements showed that there was no difference of the unbiased hit rate ( $Hu$ ) between sketch and photographed faces for all emotions. The overall unbiased hit rate is not only above the chance level (.20) but also above the conservative level of .75. Although the sketch stimuli are impoverished regarding the presence of expressive information, the similar obtained values for the  $Hu$  in comparison to the photographed face counterpart shows that this does not affect the emotion recognition. Interestingly, we found a similar pattern for the unbiased hit rate as in the validation study of the previous chapter. First of all, the happy expressions were recognized more accurate and rapid than the other emotions. Surprisingly the accuracy was lower for angry and sad expressions for both types of stimuli. This evidence, in turn, means that the initially observed lower values in the validation study were not caused by the reduced configural information on sketch faces because also the originally photographed faces leading to the same lower  $Hu$ .

The analysis of the reaction times, revealed first differences between the sketch and photographed faces. In general, the reaction time for the recognition of angry, sad and fearful expression was longer in comparison to the happy expression. Interestingly, the data showed that sketch faces with a neutral expression have led to significant longer reactions times than the happy expression. This was not observed for the photographed faces as the reaction times were similar between these two emotions. Despite the longer reaction time the effect was not observed in the unbiased hit rate. The longer reaction times while achieving similar results for the unbiased hit rate could indicate a difference in the processing of the sketch faces expressing especially a neutral expression.

The analysis of the reaction times revealed first differences between the sketch and photographed faces. In general, the reaction time for the recognition of angry, sad and fearful expression was longer in comparison to the happy expression. Interestingly, the data showed that sketch faces with a neutral expression have led to significantly longer reactions times than the happy expression. This finding was not observed for the photographed faces as the reaction times were similar between these two emotions. Despite the longer reaction time, the effect was not observed in the unbiased hit rate. The longer reaction times while achieving similar results for the unbiased hit rate could indicate a difference in the processing of the sketch faces expressing especially a neutral expression.

In order to understand the differences observed for the reaction time, eye-tracking is beneficial to gain more insights into the exploration strategy of the participants when recognizing the emotion of a sketch or photographed face stimulus. In focus of this analysis were the fixation location and duration of the participants for both stimuli types. At a first view, there was no difference of the mean fixation duration and location between sketch and photographed faces. As expected the majority of fixations were located on the diagnostic features, somehow explaining why sketch faces resulted in a similar performance in the emotion recognition task in comparison to the photographed faces. This result was in line with the study of [Schurgin et al. \(2014\)](#) who divided a face into 21 regions and found that 88.3 % of fixation were located on the diagnostic features such as eyes, nose, and mouth. In turn, the mean fixation did not vary with the stimulus type but rather depends on the facial expressions which are according to the current result of the unbiased hit rate.

In order to better understand the decoding strategy of emotional facial information between sketch and photographed faces at the early visual processing, the analysis of first and second fixations can deliver valuable details as shown for example also by other studies [Bombari et al. \(2013\)](#); [Vaidya et al. \(2014\)](#); [Eisenbarth and Alpers \(2011\)](#). The analysis of the first fixation showed that the smiling mouth of the happy expression captured immediate visual attention, whereas for the other expressions the center of the face was fixated, or an even fixation pattern across the diagnostic features was observed. This result was found for both stimulus types and the exploration

strategy found for the emotions apart from happy is linked to configural processing ([Bombari et al., 2013](#)). The smiling mouth of the happy expression has also shown in other studies to be a distinctive and unique facial feature on the happy face, resulting in quicker and more accurate performance ([Leppänen et al., 2004](#); [Calder and Young, 2005](#)).

The analysis of the second fixation revealed differences between the stimulus types. Photographed faces showed a fixation pattern similar to the first fixation including a more even fixation distribution or fixating on the center of the face for all the emotions, which implies a preferred configural processing. Conversely, for sketch faces, where angry, sad, and happy expressions induced more part-based processing due to an attentional focus on the eye region. The change in the perception strategy can be explained by the presentation of the diagnostic facial feature and removed textural information. The emphasis on the edge and contour of facial features with rich details are characteristics, which can also be found for high-spatial frequency faces. Studies with these stimuli have shown to lead to a strong attention on the eye-region for obtaining visual information to discriminate facial expression [Adolphs et al. \(2005\)](#); [Henderson et al. \(2005\)](#).

The highlight of this study is that participants processed sketch similarly and photographed faces regarding the behavioral performance, the total fixation duration, and the first fixation, but showed difference during the second fixation. During the second fixation, participants utilized configural processing for photographed faces across the facial expressions, whereas part-based processing was used for sketch faces expressing angry, happy and sad emotions. This shows the importance of a detailed analysis of the first and second fixation in order to understand the complex decoding strategy of facial perception.

### 3.3 *Publication 1*

# **Reduced configural information modulates fixation patterns but does not affect emotion recognition**

Yang Yu-Fang<sup>a,b,\*</sup>, Amorim Michel-Ange<sup>a,b</sup>, & Brunet-Gouet Eric<sup>c,d</sup>

*<sup>a</sup>CIAMS, Univ. Paris-Sud, Université Paris-Saclay, France ; <sup>b</sup>CIAMS, Université d'Orléans, France ; <sup>c</sup>HANDIRESP, UVSQ, Université Paris-Saclay, Versailles, France; <sup>d</sup>Centre Hospitalier de Versailles, Versailles, France*

Correspondence address\* CIAMS, UFR STAPS, Université Paris-Sud, 91405 Orsay Cedex, FRANCE

Corresponding author\*: yu-fang.yang@u-psud.fr

Correspondence email address: yu-fang.yang@u-psud.fr

Correspondence telephone number: + 33 1 69 15 31 22

ORCID of the corresponding author: 0000-0001-9089-6020

Funding sources: This work was supported by a PhD grant from Taiwan and Université Paris-Sud.

## **Reduced configural information modulates fixation patterns but does not affect emotion recognition**

To address the role of diagnostic features and configural information in emotion recognition, we compared fixation patterns (on eyes, nose, mouth) and recognition performance from sketched (without head contours) and photographed face counterparts. Although sketch faces supposedly induce less configural processing than photographed faces, when they convey relevant diagnostic features, recognition performance is equivalent. First fixation patterns depended on emotion. Happy mouth was the only feature that received more fixations than eyes and nose. Fixations on diagnostic features varied with stimulus type and emotion during the second fixation only. Sadness, happiness, and anger generated more fixations on eyes for sketches, suggesting a part-based perceptual strategy. Conversely, longer central fixations on photographed faces suggested more configural processing. Removal of configural information (sketched faces) did not affect emotion recognition performance, supposedly because participants used a different visual scanning strategy of part-based processing towards the eyes to compensate for the impoverished configuration.

*Keywords:* configural processing; part-based processing; eye-tracking; emotion recognition; face processing

## **Introduction**

The human face conveys a wide variety of visual information about an individual (e.g., sex, age, identity and emotional state) through a homogeneous (across humankind) set of visual features (i.e., eyes, nose, mouth, eyebrows) and their spatial relationship. Theoretical and experimental neuroscientific contributions have made it possible to define perceptual mechanisms and characterize their involvement in facial emotion recognition. Besides, the development of tools for eye tracking has opened up access to the perceptual strategies employed by the individuals. The work presented here assumes that perceptual strategies are partially influenced by emotional face stimuli according to their nature (natural or artificial) or category and that this reflects the contribution of underlying neurocognitive processes.

### ***Multiple Neurocognitive Perceptual Processes Are Identified to Recognize Facial Emotions***

The perception of faces, whether it concerns invariable (age, gender, identity) or variable (expression) aspects, can be described using concepts rooted in the neuroscience of perception research. This discipline has led to the emergence of a fairly simple distinction between mechanisms relating to the processing of local information and more integrative ones, referred to as holistic and configural processes. More precisely, configural processing refers to the simultaneous integration of all available information on the face (Le Grand, Mondloch, Maurer, & Brent, 2001; McKone, 2004; Schwarzer, Zauner, & Jovanovic, 2007), whereas part-based processing relies on the extraction of expressive information from specific parts of the face (Palermo & Rhodes, 2002; Tanaka & Farah, 1993). Numerous experiments using behavioural or electrophysiological measures have demonstrated the impact of stimuli manipulation that facilitate one or other of these processes, for instance by distorting configural processing with face-inversion paradigm. Concerning emotion, recent studies have suggested that recognition involves both types of processing (Fiorentini et al.,

2012; Meaux & Vuilleumier, 2016). Finally, the continuous cooperation of these processes is not anecdotal because it has a functional impact on the individual's social life. The importance of integrating facial features into a face configuration in order to recognize emotions successfully is obvious when comparing the performance of healthy subjects with that of patients with schizophrenia or autism spectrum disorder (Bediou et al., 2005; Kohler et al., 2003). Contrary to healthy subjects, patients with schizophrenia exhibit deficits in integrating salient features due to an inaccurate allocation of their attention on single features and less informative regions (Lee, Gosselin, Wynn, & Green, 2011; Loughland, Williams, & Gordon, 2002).

In the present work, we have been interested in the experimental paradigms used to influence perceptual processes by altering local visual information without changing face configuration. Indeed, procedures based on the disruption of configural information (i.e., scrambled faces) make it challenging to compare eye exploration strategies as the features displayed are located unexpectedly and unpredictably. More importantly, performance levels may be dramatically reduced when unusually oriented stimuli are presented, resulting in difficulties in matching experimental conditions. In line with our goals, it has been shown that the removal of facial texture information (e.g., colour, shadow, etc.) from photographed face stimuli, as in sketched faces, may dampen configural processing. In general, such investigations are based on face identification paradigms and require the comparison of photographed faces with sketched faces, or non-natural stimuli such as Arcimboldo faces. From a methodological point of view, this is particularly interesting because the design of perfectly equivalent visual stimuli benefits from image processing tools offering low-level feature control.

Some behavioural studies demonstrated that participants identified sketched faces more accurately and quickly than natural photographed faces when they were required to

identify the difference in local facial features, such as the composite face task (Zhao, Bülthoff, & Bülthoff, 2016); or the feature change detection task (Meinhardt-Injac, Persike, & Meinhardt, 2013). However, participants performed better using photographed faces rather than sketched faces when they were required to identify the whole face as a face (Meinhardt-Injac et al., 2013; Tzschaschel, Persike, & Meinhardt-Injac, 2014). In addition, an event-related potential (ERP) study found that the N170 component (a face-sensitive component peaking around 170 ms after stimulus onset over occipitotemporal areas) is larger when processing sketched faces in comparison to photographed faces, suggesting an increased difficulty in the configural processing of sketched faces (Sagiv & Bentin, 2001). Some other studies concluded that sketched faces exhibit more part-based processing due to the lack of facial texture information (as in photographed faces), which proves to be crucial surface information for features binding into a configuration (Leder, 1996; Meinhardt-Injac et al., 2013). Taking these factors into consideration, the results reported here support the idea that sketched faces reduce configural processing without modifying configural information. Comparing sketched faces with photographed faces might help to understand the importance of configural processing in emotion recognition.

### ***Several visual exploration strategies are used to recognize facial emotions***

As stated in the preamble, the advent of cognitive oculometry opens up an unprecedented avenue for analysing strategies to explore visual scenes. A first experimental fact is that human observers tend to focus their gaze on internal facial features (i.e., diagnostic features) when participants are engaged in an emotion categorization task (Gosselin & Schyns, 2001; Smith, Cottrell, Gosselin, & Schyns, 2005; Neath & Itier, 2014; Schurgin et al., 2014). However, the visual exploration of an emotional face does not follow a fixed pattern, but appears to be quickly modulated by the emotional content of the image. The most immediate

evidence is obtained by comparing fixation duration on different diagnostic features, from one emotion to another. Scheller et al. (2012) found that participants fixated more on the eye region in the recognition of fearful or neutral faces and paid more attention to the mouth of happy faces irrespective of task demands and spatial locations of face stimuli. These findings have recently been confirmed by dividing diagnostic features into three identical horizontal rectangles (eyes, nose, mouth) and by distorting the canonical face configuration on fearful, neutral and happy expressions (Elsherif, Saban, & Rotshtein, 2017). They found that participants recognized fear from the eyes and happiness from the mouth more easily. However, participants were more likely to identify the presence of fearful eyes when the fearful mouth was also presented.

Most studies using an emotion categorization task have demonstrated the preference for a region for specific facial expressions. For example, participants tend to look at the eye region when trying to categorize fearful faces (Smith et al., 2005), or sad faces (Eisenbarth & Alpers, 2011), but look at the mouth region in order to recognize happy faces (Calvo & Nummenmaa, 2008; Calvo, Nummenmaa, & Avero, 2010). However, facial expressions might not always be recognized merely by a single facial feature, especially when the task contains more than two facial expressions. For example, participants tend to gather configural information by fixating on the centre of the face when distinguishing fearful expressions from other facial expressions (Bombari et al., 2013; Smith et al., 2005). Taken together, these results demonstrate that human observers employ different perceptual mechanisms that not only rely on isolated cues but also depend on global configuration to capture the relevant emotional information.

As well as refining the description of exploration strategies, other results made it possible to question the perceptual processes at work. Bombari et al. (2013) used intact, blurred and scrambled faces as stimuli and demonstrated that different face processing

approaches (configural and part-based processing) are employed, depending on the facial expression displayed. They argued that when the fixation is located over the centre of the face (i.e., on the nose region), this can be considered as an indicator for configural processing. This suggestion is supported by a study of Hsiao and Cottrell (2008) who demonstrated that two fixations are sufficient to identify a neutral facial expression. They observed that fixations are directed upon the centre of the face regardless of the spatial location of the stimuli on the screen, which reflects configural processing. Schwarzer, Huber, & Dümmler (2005) investigated the optimal gaze location for face processing and suggested that participants not only fixate on one facial feature but also locate their fixations on the upper part of faces slightly below the eyes, known as the location of the centre of the face. Note that the centre of the face refers to the mass of the face rather than to the geometric centre of the face (Rossion, 2014). Supporting evidence came from patients with prosopagnosia who have difficulties in processing configural face information and therefore do not fixate the centre of the face (i.e., nasion area) in comparison to healthy subjects (Orban de Xivry, Missal, & Lefevre, 2008; Peterson & Eckstein, 2012).

If facial stimuli influence exploration strategies, the question is: when does this influence occur? The analysis of the successive fixations patterns may provide interesting elements in order to understand the cognitive processes that are employed and their temporal sequence. A few studies have proposed a more precise examination of eye-tracking data that considers the temporal order and duration of fixations (see Table 1). Concerning the first fixation, some authors showed that fixations were oriented towards the centre of the face whatever the emotion category (Schurgin et al., 2014), a fact that could denote the use of an initial holistic or configural processing. Conversely, other authors suggested that the fixation pattern is influenced by the emotion category from the first fixation (Bombari et al., 2013; Vaidya, Jin, & Fellows, 2014). In other words, participants search for the distinctive

diagnostic feature as a cue for initial emotion identification, such as the mouth for expressions of happiness or fear, or the eye for recognizing sad expressions. Note that the discrepancies concerning the first fixation may also result from the ambiguous definition of the first fixation. The very first fixation following the facial stimulus onset was discarded in the studies of Bombari et al. (2013) and Eisenbarth & Alpers (2011) from Table 1. However, some authors did not define the first fixation explicitly (Schurgin et al., 2014; Vaidya et al., 2014). In order to interpret the results of the temporal fixation organization when categorizing emotions, the fixations should be compared in the same temporal order. The following fixation (i.e., the second one) is somewhat more oriented towards the eyes than other regions in comparison with the first fixation across emotions (Eisenbarth & Alpers, 2011). However, other research found that the nose region is fixated longer during the second fixation, or even no particular region of the face at all (Bombari et al., 2013). Thus, it remains unclear whether the second fixation reflects preferential part-based strategies related to some emotional contents. From a methodological point of view, one can conclude that investigating the fixation location and duration of the first and second fixations towards the areas of interest (AOIs) is crucial to understand how facial expressions are decoded. Another finding is that there is no systematic temporal pattern of fixations across AOIs for any given emotion, at least with photographed face stimuli, making it more difficult to understand the relative contribution of part-based and configural processing

### ***Objectives***

In the present study, we investigated the relative contribution and the temporal characteristics of part-based and configural processing during emotion recognition by comparing natural photographs and impoverished visual stimuli. We therefore designed a new set of sketch stimuli that retains diagnostic features only, with no facial texture or non-informative parts by

transforming photographed faces from the Radboud Faces Database (RaFD, Langner et al., 2010), which contains five basic facial expressions (angry, sad, fearful, happy and neutral). We directly compared the sketched stimuli with original photographs in an emotion recognition task using both behavioural and eye-tracking measures. With the analysis of fixation-sequences, we investigated the similarities and differences in the early face processing between the two facial stimuli and also between the emotions. Within the framework of our working model, derived from the literature described above, we formulated several hypotheses.

As a preliminary assumption, a set of sketched stimuli might reach accuracy and reaction times comparable to those of natural photograph counterparts, because sketches contain the relevant information in the diagnostic feature regions. The emotion category should elicit some differences in the exploration pattern: the mouth should be fixated in happy expressions and the eye region will be targeted when recognizing anger and sad expressions. We hypothesized that participants might employ configural processing, such as fixating on the centre of the face or equally-distributed fixations between features, to gather global information for distinguishing one emotion from another. Concerning the sequences of eye fixations, we hypothesized that all the stimuli will generate a first central fixation, that would reflect some inaugural, holistic/configural perceptual input. As sketched stimuli are thought to favour local processing, we expect longer subsequent fixations on the mouth or the eyes, reflecting the enhancement of this perceptual pathway.

## **Experiment**

We created a new set of sketched faces composed of isolated diagnostic features (i.e., eyes, nose and mouth) expressing five basic emotions (i.e., angry, sadness, happiness, fear and neutrality) to study the temporal organization of fixations towards the diagnostic features

within sketches and their photographed face counterparts. This sketched stimulus set was created from photographed faces taken from the Radboud faces database ((Langner et al., 2010).

This set of sketched faces being new, we aimed to select only sketch stimuli for each emotion that would lead to equivalent high-level behavioural recognition performance. Therefore, we first conducted a preliminary behavioural experiment to validate the sketched faces and then selected the ten best-recognized sketched faces for each emotion (see Online Resource). The selected sketched faces were used in the present eye-tracking study in which the fixation location and duration were examined and compared to the photographed face counterparts.

## ***Methods***

### *participants*

40 healthy subjects participated in this study (a new group that was not in the preliminary study) and were randomly assigned to one of two “stimulus types” group: sketched faces (mean age 26.7 yrs, SD= 3.9, age range: 19-33) or photographed faces (mean age 26.4 yrs, SD= 4.3, age range: 19-33). Each group had 20 participants (10 males and 10 females). All participants were right-handed and had normal or corrected-to-normal vision and gave their prior informed consent in accordance with French regulations and the Declaration of Helsinki. One subject was excluded before the acquisition of behavioural data due to the failure of initial eye movement calibration.

### *apparatus and stimuli*

Stimuli were presented on a 17-inch monitor (resolution 1280 x 1024-pixel, screen width is 340 mm x height is 270 mm) with 60 Hz refresh rate. Eye movements were recorded using

EyeLink Remote 1000 (SR Research Ltd., Ottawa, Canada) at a 1000 Hz monocular sampling rate and the spatial resolution was better than 0.5°. E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used to present the stimuli and to record the behavioural performance. The size of the stimulus images was 400 x 400 pixels, which is equal to 10° x 10° visual angle at a view distance of 60 cm.

The stimuli were 50 images of sketched faces and 50 images of photographed faces. The sketched faces were transformed from the RaFD and selected from the preliminary experiment (see Online Resource for details). The photographed face counterparts were therefore chosen to be comparable with sketched faces. Each set of 50 facial expressions corresponded to a combination of ten facial identities (five females and five males) with five emotions (neutral, fearful, angry, sad, and happy expressions). Depending on the group of participants, these 50 stimuli were either sketched or photographed face stimuli. Each participant experienced only one type of facial stimulus during the experiment. The size of the photographed stimuli was reduced to match that of the sketched facial stimuli (400 x 400 pixels).

### *procedure*

The participants were seated in front of the computer screen with the eye-tracker camera positioned below the screen. The eye tracking system was calibrated for each participant using a nine-point calibration and validation procedure. The participants were seated at a distance of 60 cm between their head and the display screen. A head support system was used to stabilize each participant's forehead and chin during recording. Once participant visual angles were accurate to under 0.5° for calibration and validation, drift correction was used to measure any significant difference between the computed fixation position and the target.

The experiment involved both a practice and experimental session. The practice sessions were subdivided into a response key learning session (session 1) and an emotion recognition session (with sketched face or photographed face stimuli, depending on the group; session 2). Each practice session involved response feedback. To continue with the experimental session, the participants were required to reach 100% accuracy in session 1, and 75% accuracy in session 2. Once participants had passed the practice sessions, the procedure for the experimental session was the same as in practice session 2, except that the stimuli from the practice session were not included, and the participants were not provided with a feedback on their response accuracy. For details on the practice session, see Online Resource.

The emotion recognition sessions were forced-choice tasks in which the participants were required to indicate which emotion was displayed out of five possible responses: neutrality, anger, fear, happiness and sadness. Each session comprised 50 randomized sketched faces or photographed faces, depending on the group. Each trial began with a 500 ms fixation cross, followed by the facial stimulus, which remained displayed on the screen until the participant responded (self-paced). A 500 ms blank screen followed this. The task was self-paced and had no time limit. The participants were instructed to respond as quickly and accurately as possible. They were randomly assigned to either recognize the emotional state of sketched faces or photographed facial stimuli using a between-participants design to avoid carry-over effects.

### *Data Analyses*

behavioural data. Response accuracy and reaction time (RT) for trials with correct responses were analysed for each stimulus. Unbiased hit rates (Hu) were computed as a measure for accuracy for each emotional state category. The use of Hu was introduced by Wagner (1993) and is the most conservative measure of accuracy in a categorical judgment study. Hu can

vary between 0 and 1. The unbiased hit rate reveals whether the accuracy increase was due to bias (i.e., chance level of guessing) or to increased sensitivity of target emotion labelling.

Thereby the  $H_u$  considers both the hit and false alarm rate. The hit rate describes the proportion of correctly categorized target emotion trials, whereas the number of false alarms is the number of misclassified target emotion trials. The calculation of  $H_u$  is based on a confusion matrix and can be expressed in the following equation (Wagner, 1993).

$$H_u = Hits^2 / (A_i B_i) \quad (1)$$

In the above equation (1) Hits is the frequency of correct answer trials,  $A_i$  the number of trials per response condition and  $B_i$  the number of trials per stimulus with  $i$  as the target emotion. Then,  $H_u$  is further normalized using an arcsine transformation, the value varying between 0 and 1. Additionally, Bonferroni-corrected  $t$ -tests were conducted to test our conservative criterion ( $H_u = .75$ ) and chance level of response ( $H_u = .20$ ) for each emotion.

Analyses of the RTs were based on correct trials. We excluded trials with RTs greater than the value of “mean + 3SD” from our analysis to reduce the influence of outliers, even with only 5.4% of the total trials. These trials were also excluded from our unbiased hit rate analysis. The RT for correct answers and the unbiased hit rate ( $H_u$ ) was further analysed by repeated measures ANOVA, with stimuli (either sketched or photographed faces) as a between-subject factor, and emotion (at five levels) as a within-subject factor. Post-hoc Tukey’s test (HSD) analyses were performed when the main effects or interactions were significant. Greenhouse-Geisser corrections were applied when the sphericity assumption was violated ( $\epsilon$  values are provided in this case).

*eye-tracking data.* Analyses were performed only on correct response trials. Prior to eye movement analysis, AOIs were defined such as the eye region (the left- and right-eye region, including eyebrows were grouped as a single “eye” AOI), nose, and mouth region using

EyeLink Dataviewer software (SR Research Ltd., Ottawa, Canada) for each emotion category and stimulus type, see Figure 1. The parameters defining the eye movement were computed by EyeLink 1000 (SR Research Ltd., Ottawa, Canada). Fixation is detected with velocity, acceleration, and motion thresholds of  $30^\circ/\text{sec}$ ,  $8,000^\circ/\text{sec}^2$ , and  $0.15^\circ$ , respectively.

We (a) generated heat maps that visualize the sum of fixation durations and location during an entire presentation of stimuli per emotion, and (b) analysed the mean fixation duration as well as the first and second fixation on AOIs. Heat maps were provided to visualize the spatial distribution of fixation density alongside the results of the mean fixation duration. They were generated by using Matlab 8.4 Image processing toolbox (The Mathworks Inc., Natick, MA), with a Gaussian kernel density estimation (Jones, Marron, & Sheather, 1996). Here, the amplitude and distribution spread were proportional to the fixation duration (with  $k=1$ ; for the exact formula, see Nikolaev, Jurica, Nakatani, Plomp, & van Leeuwen, 2013).

We analysed both the mean fixation duration as well as the first and second fixation per emotion for each AOI during an entire stimulus presentation. The mean fixation duration was calculated by dividing the sum of the fixation duration of each AOI by the number of given valid trials in a condition. The location and duration of the first and second fixations on AOIs were analysed. Noted that the very first fixation following the stimulus onset was discarded because the participant's fixations remained on the centre of the screen (influenced by the location of the precedent fixation cross). Statistical analyses were conducted on the fixation duration for each AOI. Mixed-model ANOVAs were conducted with five emotions (anger, fear, happiness, sadness, and neutrality) and three AOIs (eyes, nose, and mouth) as within-subject factors. The stimulus type (sketched or photographed faces) was treated as a between-subjects factor. Post-hoc Tukey HSD tests were conducted for significant main effect and interactions. Greenhouse-Geisser corrections were applied when the sphericity

assumption was violated ( $\epsilon$  values are provided in this case). Data processing and statistical analyses were performed using Statistica (StatSoft, Inc).

## **Results**

### *behavioural result*

A repeated-measures ANOVA on unbiased hit rates (Hu) revealed no main effect of stimulus type ( $F(1, 38) = .39$ , n.s.) and showed a main effect for emotion ( $F(4, 152) = 13.93$ ,  $\epsilon = 0.60$ ,  $p < .0001$ ,  $\eta^2_p = .27$ ). Post-hoc comparisons showed that the Hu for happiness (.98) was higher than for the other emotions,  $ps < .01$ . Hu for sadness was lower than for fearful, happy and neutral expressions ( $ps < .01$ ) (see Figure2). Bonferroni-corrected  $t$ -tests (with  $\alpha = .05/5$ ) showed that the recognition rate Hu for each emotion was greater than the .75 conservative criterion, with all  $p$ -values being less than .01. The ANOVA showed no interaction between the stimulus type and emotion ( $F(4, 152) < 1$ , n.s.).

The repeated measures ANOVA on mean RTs showed no main effect of stimulus type ( $F(1, 38) < 1$ , n.s.), but did show a significant main effect of emotion ( $F(4, 152) = 21.99$ ,  $\epsilon = 0.71$ ,  $p < .0001$ ,  $\eta^2_p = .37$ ). Post-hoc tests showed that mean RT was shorter for happiness than for any other emotion ( $ps < .0001$ ). RTs were longer for angry faces compared with those for happy and neutral expressions ( $ps < .01$ ). However, the effect of emotion varied significantly with the stimulus type ( $F(4, 152) = 3.41$ ,  $\epsilon = 0.71$ ,  $p < .05$ ,  $\eta^2_p = .08$ ). Post-hoc tests identified a mean RT advantage for happiness on sketched faces ( $ps < .001$ ), but not for photographed faces, because there was no happiness advantage compared with neutral expressions ( $p > .5$ ) (see Figure3).

### *eye-tracking results*

*mean fixation duration.* Heat maps, showing the sum fixation durations for all presented stimuli, revealed that most fixations were located on the inner facial features for both photographs and sketch stimuli. These fixations varied depending on the type of emotion. ANOVA on the mean fixation duration revealed a significant Emotion x AOIs interaction ( $F(8, 304) = 16.29, \epsilon = 0.54, p < .0001, \eta^2p = .30$ ) that did not vary with the stimulus type ( $F(8, 304) = 1.75, \epsilon = 0.54, p = .087, \eta^2p = .044$ , see Figure 4). Post-hoc analyses on the former interaction showed different mean fixation duration patterns for AOIs, depending on the emotion. The participants spent significantly less time fixating on the mouth region than on the eyes and nose for angry and sad faces ( $ps < .001$ ). For fearful faces, mean fixation durations varied significantly between each AOI ( $ps < .01$ ), with mean fixation durations for nose > eye > mouth. In contrast, mean fixation durations did not vary with AOI for happy faces ( $ps > .88$ ). In addition, mean fixation durations for the mouth on the happy expressions were not significantly different from those for the mouth for other emotions ( $ps > .11$ ). For neutral faces, the participants fixated significantly more on the eyes than on the mouth ( $p < .001$ ), but mean fixation durations did not significantly differ between the mouth and the nose ( $p = .078$ ).

The relative importance of each AOI was also examined across emotions. Mean fixation durations for the nose were greater for fearful faces than for the other emotional expressions ( $ps < .04$ ). Mean fixation durations on the nose for angry faces were greater than for neutral, sad and happy faces ( $ps < .04$ ). In contrast, mean fixation durations on the nose did not differ for sad, happy and neutral faces ( $ps > .75$ ). The eyes received similar mean fixation durations for anger, fear and sadness ( $ps > .16$ ). The mean fixation durations for the eyes were significantly shorter for happy faces than for the other emotions ( $ps < .01$ ), except for neutral faces ( $p = .05$ ). Fearful faces showed greater mean fixation durations for the

mouth than for the other emotions ( $ps < .01$ ), except for happy faces ( $p = .98$ ). Neutral, angry and sad faces received a similar mean fixation duration for the mouth ( $ps > .99$ ).

*first fixation duration and location.* The ANOVA on the first fixation duration revealed a significant interaction between emotion and AOI ( $F(8, 304) = 9.85, \varepsilon = .65, p < .0001, \eta^2_p = .21$ ). However, this interaction did not vary with the stimulus type ( $F(8, 304) < 1, n.s.$ ). Heatmaps showed that first fixation locations were the same AOI per emotion between sketched and photographed faces (see Figure 5A). The significant interaction of Emotion x AOI showed that first fixation duration patterns varied with emotion. Follow-up post-hoc analyses showed that the first fixation duration was equivalent across AOI for neutral and sad faces ( $ps > .99$ ). However, participants fixated significantly less on the mouth than on the nose region ( $p < .002$ ) for angry expressions, but for a similar duration around the eyes and nose (see Figure 6A). In contrast, there were significantly longer fixations on the mouth than on the eyes or nose ( $ps < .0001$ ) for happy faces. For fearful expressions, the participants fixated for a significantly shorter time on the eyes than on the nose or mouth ( $ps < .02$ ), and the heat map showed that fixations landed more frequently below the nostrils and on the left eye.

Follow-up post-hoc comparisons examined each AOI across emotions. The participants fixated the mouth region for similar amounts of time ( $ps > .92$ ) on happy and fearful faces, but fixated significantly more time on happy and fearful faces than for all other emotions ( $ps < .0001$ ). They focused on the eye region more for angry than happy faces ( $p < .04$ ), but in a comparable way for other emotions. Participants fixated more on the nose for fearful faces than for neutral, sad and happy faces ( $ps < .005$ ).

*second fixation duration and location.* The ANOVA analysis on the second fixation duration showed a significant interaction between emotion and AOI ( $F(8, 304) = 2.87, \varepsilon = .69, p = .01$ ,

$\eta^2_p = .07$ ). In addition, this interaction varied significantly with the stimulus type ( $F(8, 304) = 4.54, \varepsilon = .69, p < .0001, \eta^2_p = .11$ ). the heatmap of the sum of the fixation duration also revealed different fixation patterns, see Figure 5B. Here, we detailed only significant AOI x stimulus type interaction for each emotion, for the sake of simplicity. The ANOVAs showed non-significant interactions for angry faces ( $F(2,76) = 1.80, p = .17$ ) and neutral faces ( $F(2,76) < 1, n.s.$ ). In contrast, the ANOVAs revealed significant interaction for faces showing fear ( $F(2,76) = 5.01, p < .01$ ), happiness ( $F(2,76) = 7.1, p < .002$ ) and sadness ( $F(2,76) = 4.7, p = .01$ ). Post-hoc comparisons for fearful expressions showed that the participants spent more time fixating on the nose of photographed faces than sketched faces ( $p = .003$ ). Second fixation durations for the eye region were significantly greater for sketched faces than for their photographed face counterparts when recognizing expressions of happiness ( $p < .01$ ) and sadness ( $p < .01$ ), see Figure 6B. Additional post-hoc comparisons revealed greater durations for the eye region than for the nose and the mouth areas for angry, happy and sad sketched faces than for photographed faces ( $ps < .05$ ).

## Discussion

Recognizing facial expressions is essential for social communication and has been studied extensively. The location and duration of eye fixations can unveil the visual scanning strategy employed for decoding facial expressions. The aim of this study was to examine, through a step-by-step analysis of early perceptual strategies, the role of face processing (e.g., part-based and configural processing) in facial emotion recognition and whether participants extract more featural information from the newly created sketched faces in comparison to their photographed face counterparts.

In agreement with the preliminary hypothesis, the results show that the participants have a high performance regarding the correct emotion recognition, independent of the stimulus

type. This is noticeable as these stimuli show notable differences concerning the presence of visual information. However, our study shows that the sketched facial features convey sufficient expression to enable an accurate emotion recognition level similar to photographed faces. The importance of diagnostic features for emotion recognition has also been shown in other studies and supports the present results (Schurgin et al., 2014; Smith et al., 2005). To the best of our knowledge, this is the first time that a direct comparison between a sketched face and visually matched photographed counterparts in an emotion recognition task has been performed and similar behavioural performance obtained. Across all of the trials for both sketched and photographed faces, the eye fixations are directed to the diagnostic features (the eyes, nose, mouth) where normally relevant emotional information can normally be found. In the present study, we found that neither the mean fixation duration and location nor the behavioural performance changed much regarding the stimulus type used.

In our assumptions, we referred to the simple distinction between configural/holistic and part-based processing. Indeed, we considered a certain number of knowledge elements that distinguish these treatments in both in terms of brain mechanisms and of the influence of experimental conditions. In addition, we followed the suggestion that visual exploration strategies provide information about underlying processes. For instance, in a face identification study using neutral photographed faces, Hsiao and Cottrell (2008) hypothesized that a face could be recognized in only two fixations. As the first and second fixation were located on the nose, they linked these findings to configural processing. Piepers and Robbins (2012) suggested that configural processing also occurs when fixations are equally distributed among facial features because participants show no preference to any particular feature but rather integrate information from all parts of the face. In the same line, it is interesting to discuss manifestations of global processing and to see when they occur in the context of emotion recognition.

The analysis of the first fixation provides some contrasting results demonstrating an early impact of the emotion category on the scanning patterns. Firstly, the participants' preferred fixation locations were on the centre of the face for the angry and fearful expressions, a fact that was interpreted as the presence of configural processing. This finding is in line with a study of Bombari et al. (2013) in which they suggested that angry expressions require more configural processing than happy expressions. Secondly, when other expressions such as anger or sadness are presented, the lack of salient features results in a broader distribution of the fixations over the facial features and leads to a poorer performance in emotion recognition. The distribution of fixation over different facial features with similar durations may result from the fact that these facial expressions are unrecognizable from a single facial feature. Taken together these results may be interpreted as demonstrating the importance of the inaugural holistic/configural strategy of looking at the centre of the face when highly salient emotional stimuli are presented. Less salient stimuli, such as neutral or sad ones, lead to less contrasted patterns.

Interestingly, the analysis of the happy emotion stimuli brings evidence to the fact that the content influences the scanning strategy from the very first fixation and informs us about the underlying processing. The smiling mouth of the happy expression received considerable attention from the participants, suggesting that part-based processing occurs early once a distinctive feature emerges. Similarly to other studies (Calvo & Nummenmaa, 2008; Calvo et al., 2010; Eisenbarth & Alpers, 2011), the smiling mouth of the happy expression leads to a faster and more accurate recognition. In comparison to the other emotions, there is no other facial feature which captures a fixation duration as long as the smiling mouth of happy expressions during the first fixation. This unique feature enables participants to recognize a happy expression in peripheral vision even when the mouth region only is presented (Calvo, Fernández-Martín, & Nummenmaa, 2014). Evidence also shows that the high saliency of the

smiling mouth leads to higher accuracy and faster processing times in emotion categorization.

The second fixation also provides information on the underlying perceptual processes, considering the information already obtained during the first step.

In this study, the fact that no preferential AOI was fixated during the second fixation for the photographed stimuli, suggests that participants tend to employ configural processing. However, the use of sketched faces as stimuli revealed a preferred fixation location on the eye region. This distinctive difference in comparison to the photographed faces can be found for the angry, happy and sad expressions. While several studies have found that photographed faces evoke more configural processing than sketched faces, it is the first time that this phenomenon was observed during the second fixation. Note that the participants only viewed one stimulus type during experiments. Thus, these differences between stimuli cannot be accounted for by some artifactual effect of the contrast between conditions. The location of the second fixation on the eye region, in the group looking at sketches, for angry, happy and sad facial expressions, expresses the involvement of part-based processing in the emotion recognition. Surprisingly, our participants located their fixations on the eye region instead of the mouth region for happy expressions. This suggests that the recognition of happy expressions might be accomplished during the first fixation, whereas the second fixation is directed to the eye region for final confirmation. Using high spatial frequency face stimuli with the bubble technique (random location of visual gaussian "bubbles" were made visible on the face), Adolphs et al. (2005) found a preferred tendency to fixate the eye region for the same emotions (happy, sad and angry) among healthy participants. The newly created sketched faces share similar characteristics to high-spatial frequency faces, such as the emphasis of the edge and contour of the facial features with rich detail. But this result, along with the initial preferred fixation on the mouth, may indicate coordinated part-based

processing that orients perception strategy to the most relevant diagnostic parts. Therefore, the fixation pattern did not vary with stimulus type for the first fixation, but for the second.

Fearful expressions seem to play a special role given that they induce perceptual strategies that are more compatible with configural processing than to part-based processing. Although these expressions contain salient facial features such as wide-open eyes (Adolphs et al., 2005; Whalen, 2004), our results show that the correct fearful face categorization did not solely rely on this featural information. This assumption is supported by a study of Bombari et al. (2013), who demonstrated that the recognition of fearful faces is better when configural information is present such as in an intact face version rather than a scrambled face version. This could explain the use of configural processing for the recognition of fearful faces although important visual information is missing in comparison to the photographs. The observed differences, especially for the second fixation, need further investigation to establish the relationship between part-based and configural processing. Future investigations could use brain measures, such as electroencephalography (EEG), to confirm the potential differences with regards to face processing that cannot be captured by eye-tracking and behavioural approaches. Of specific interest would be the face-sensitive component N170, which peaks around 170 ms after the stimulus onset and is identified as an index of the balance between part-based and configural processing (Bentin, Golland, Flevakis, Robertson, & Moscovitch, 2006; Caharel et al., 2013; Rossion, 2009).

## **Conclusions**

The present study confirms that diagnostic features contribute greatly to emotion recognition and provide a start point for the decoding of facial expression processing in both sketched and photographed faces. The design of a new picture set perfectly matched to the RaFD, allowed this question to be addressed accurately. Independent of the presence of rich visual

information on photographed faces, behavioural performance and mean fixation patterns showed a similar behaviour for the newly created sketched faces in comparison to the originally photographed faces. However, analysing the early visual processing of emotional expressions gave more detailed insights into the temporal dynamics of emotion recognition and helped to reveal the presence of either configural or part-based processing. Thereby, the location of the first fixation seems to indicate more configural processing irrespective of the presented emotion and stimulus type. An exemption was the happy expression where the highly distinctive facial feature of the smiling mouth captured immediate visual attention. Differences concerning the fixation pattern and the involved face processing mode were found for the second fixation: where configural processing was still preferred for photographed faces, part-based processing was observed for sketched faces as indicated by the fixation location on the eye region. To sum up, specific fixation patterns may be due to the emotional cues provided by specific facial expressions and facial textures (e.g., sketched and photographed faces).

## References

- Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, 433(7021), 68–72.
- Beaupré, M. G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology*, 36(3), 355–370.
- Bediou, B., Franck, N., Saoud, M., Baudouin, J.-Y., Tiberghien, G., Daléry, J., & d'Amato, T. (2005). Effects of emotion and identity on facial affect processing in schizophrenia. *Psychiatry Research*, 133(2–3), 149–157.  
<https://doi.org/10.1016/j.psychres.2004.08.008>
- Bentin, S., Golland, Y., Flevaris, A., Robertson, L. C., & Moscovitch, M. (2006). Processing the trees and the forest during initial stages of face perception: Electrophysiological evidence. *Journal of Cognitive Neuroscience*, 18(8), 1406–1421.
- Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion recognition: The role of featural and configural face information. *The Quarterly Journal of Experimental Psychology*, 66(12), 2426–2442.  
<https://doi.org/10.1080/17470218.2013.789065>
- Caharel, S., Leleu, A., Bernard, C., Viggiano, M.-P., Lalonde, R., & Rebaï, M. (2013). Early holistic face-like processing of Arcimboldo paintings in the right occipito-temporal cortex: evidence from the N170 ERP component. *International Journal of Psychophysiology*, 90(2), 157–164.
- Calvo, M. G., Fernández-Martín, A., & Nummenmaa, L. (2014). Facial expression recognition in peripheral versus central vision: role of the eyes and the mouth. *Psychological Research*, 78(2), 180–195. <https://doi.org/10.1007/s00426-013-0492-x>
- Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: Salient physical features guide effective visual search. *Journal of Experimental Psychology: General*, 137(3), 471–494. <https://doi.org/10.1037/a0012771>
- Calvo, M. G., Nummenmaa, L., & Averó, P. (2010). Recognition advantage of happy faces in extrafoveal vision: Featural and affective processing. *Visual Cognition*, 18(9), 1274–1297. <https://doi.org/10.1080/13506285.2010.481867>
- Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion*, 11(4), 860–865. <https://doi.org/10.1037/a0022758>

- Elsherif, M. M., Saban, M. I., & Rotshtein, P. (2017). The perceptual saliency of fearful eyes and smiles: A signal detection study. *PloS One*, *12*(3), e0173199.
- Fiorentini, C., Schmidt, S., & Viviani, P. (2012). The Identification of Unfolding Facial Expressions. *Perception*, *41*(5), 532–555. <https://doi.org/10.1068/p7052>
- Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.
- Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106.
- Hsiao, J. H., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, *19*(10), 998–1006.
- Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., ... Gur, R. C. (2003). Facial Emotion Recognition in Schizophrenia: Intensity Effects and Error Pattern. *American Journal of Psychiatry*, *160*(10), 1768–1774. <https://doi.org/10.1176/appi.ajp.160.10.1768>
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, *24*(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception: Early visual experience and face processing. *Nature*, *410*(6831), 890.
- Leder, H. (1996). Line drawings of faces reduce configural processing. *Perception*, *25*(3), 355–366.
- Lee, J., Gosselin, F., Wynn, J. K., & Green, M. F. (2011). How Do Schizophrenia Patients Use Visual Information to Decode Facial Emotion? *Schizophrenia Bulletin*, *37*(5), 1001–1008. <https://doi.org/10.1093/schbul/sbq006>
- Loughland, C. M., Williams, L. M., & Gordon, E. (2002). Visual scanpaths to positive and negative facial emotions in an outpatient schizophrenia sample. *Schizophrenia Research*, *55*(1–2), 159–170. [https://doi.org/10.1016/S0920-9964\(01\)00186-4](https://doi.org/10.1016/S0920-9964(01)00186-4)
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet*, (1998).
- McKone, E. (2004). Isolating the special component of face recognition: peripheral identification and a Mooney face. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 181.

- Meaux, E., & Vuilleumier, P. (2016). Facing mixed emotions: Analytic and holistic perception of facial emotion expressions engages separate brain networks. *NeuroImage*, 141, 154–173. <https://doi.org/10.1016/j.neuroimage.2016.07.004>
- Meinhardt-Injac, B., Persike, M., & Meinhardt, G. (2013). Holistic Face Processing is Induced by Shape and Texture. *Perception*, 42(7), 716–732. <https://doi.org/10.1068/p7462>
- Neath, K. N., & Itier, R. J. (2014). Facial expression discrimination varies with presentation time but not with fixation on features: a backward masking study using eye-tracking. *Cognition & Emotion*, 28(1), 115–131.
- Nikolaev, A. R., Jurica, P., Nakatani, C., Plomp, G., & van Leeuwen, C. (2013). Visual encoding and fixation target selection in free viewing: presaccadic brain potentials. *Frontiers in Systems Neuroscience*, 7. <https://doi.org/10.3389/fnsys.2013.00026>
- Orban de Xivry, J. J., Missal, M., & Lefevre, P. (2008). A dynamic representation of target motion drives predictive smooth pursuit during target blanking. *Journal of Vision*, 8(15), 6–6. <https://doi.org/10.1167/8.15.6>
- Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition*, 82(3), 225–257. [https://doi.org/10.1016/S0010-0277\(01\)00160-3](https://doi.org/10.1016/S0010-0277(01)00160-3)
- Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, 109(48), E3314–E3323. <https://doi.org/10.1073/pnas.1214269109>
- Piepers, D. W., & Robbins, R. A. (2012). A Review and Clarification of the Terms “holistic,” “configural,” and “relational” in the Face Perception Literature. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00559>
- Rossion, B. (2009). Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis. *Acta Psychologica*, 132(3), 300–312. <https://doi.org/10.1016/j.actpsy.2009.08.002>
- Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends in Cognitive Sciences*, 18(6), 310–318.
- Sagiv, N., & Bentin, S. (2001). Structural encoding of human and schematic faces: holistic and part-based processes. *Journal of Cognitive Neuroscience*, 13(7), 937–951.
- Scheller, E., Büchel, C., & Gamer, M. (2012). Diagnostic features of emotional expressions are processed preferentially. *PloS One*, 7(7), e41792.

- Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision*, *14*(13), 14–14.
- Schwarzer, G., Huber, S., & Dümmler, T. (2005). Gaze behavior in analytical and holistic face processing. *Memory & Cognition*, *33*(2), 344–354.  
<https://doi.org/10.3758/BF03195322>
- Schwarzer, G., Zauner, N., & Jovanovic, B. (2007). Evidence of a shift from featural to configural face processing in infancy. *Developmental Science*, *10*(4), 452–463.  
<https://doi.org/10.1111/j.1467-7687.2007.00599.x>
- Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and Decoding Facial Expressions. *Psychological Science*, *16*(3), 184–189.  
<https://doi.org/10.1111/j.0956-7976.2005.00801.x>
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, *46*(2), 225–245.
- Tzschaschel, E., Persike, M., & Meinhardt-Injac, B. (2014). The effect of texture on face identification and configural information processing. *Psihologija*, *47*(4), 433–447.  
<https://doi.org/10.2298/PSI1404433T>
- Vaidya, A. R., Jin, C., & Fellows, L. K. (2014). Eye spy: The predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition*, *133*(2), 443–456.
- Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28.
- Whalen, P. J. (2004). Human Amygdala Responsivity to Masked Fearful Eye Whites. *Science*, *306*(5704), 2061–2061. <https://doi.org/10.1126/science.1103617>
- Zhao, M., Bülthoff, H. H., & Bülthoff, I. (2016). A shape-based account for holistic face processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(4), 584–597. <https://doi.org/10.1037/xlm0000185>

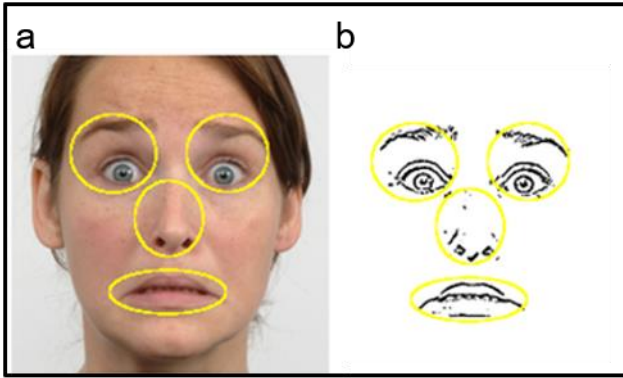


Figure 1. Example of AOIs for a fearful female expressor for (a) photographed face stimuli, and (b) sketched face stimuli.

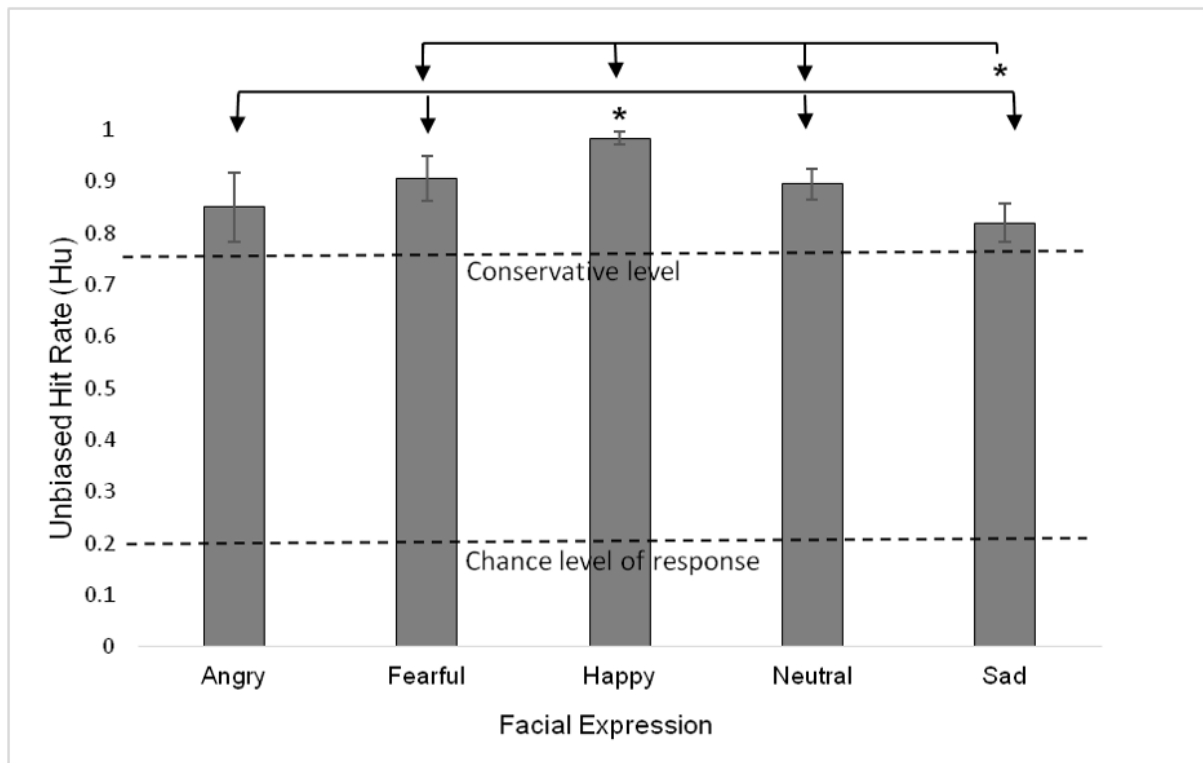


Figure 2. Mean unbiased hit rate (Hu) (with 95% Confidence Interval as error bars) per emotion. Dashed lines indicate the conservative level of  $Hu = .75$ , and chance level of response,  $Hu = .20$ .

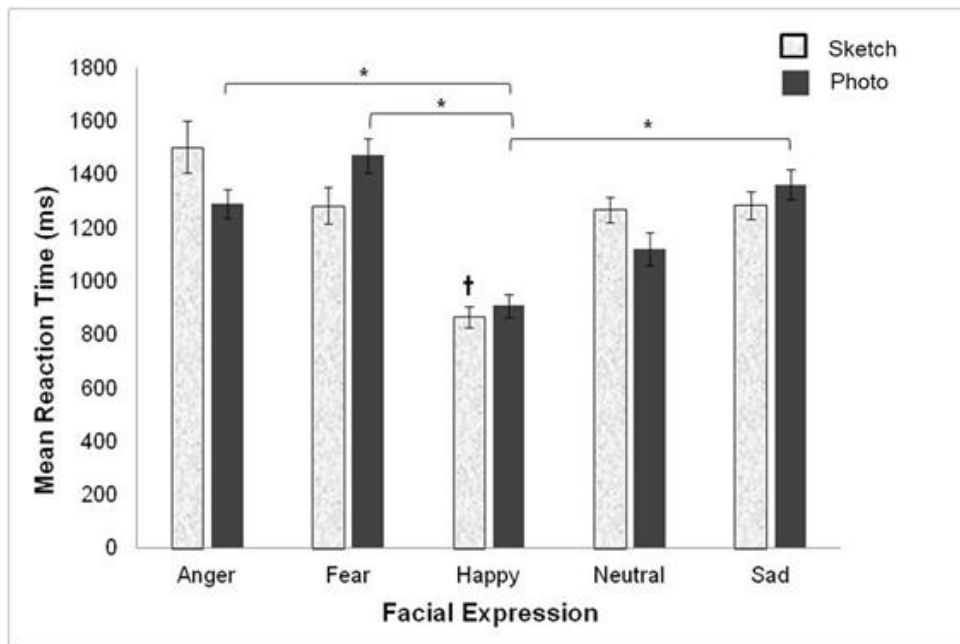


Figure3. Mean reaction time (with 95% Confidence Interval error bars) for sketch and photograph faces per emotion. \*: significant difference for photograph faces ( $p < .01$ ); † this condition differed significantly ( $p < .001$ ) from every other emotion for the sketched faces.

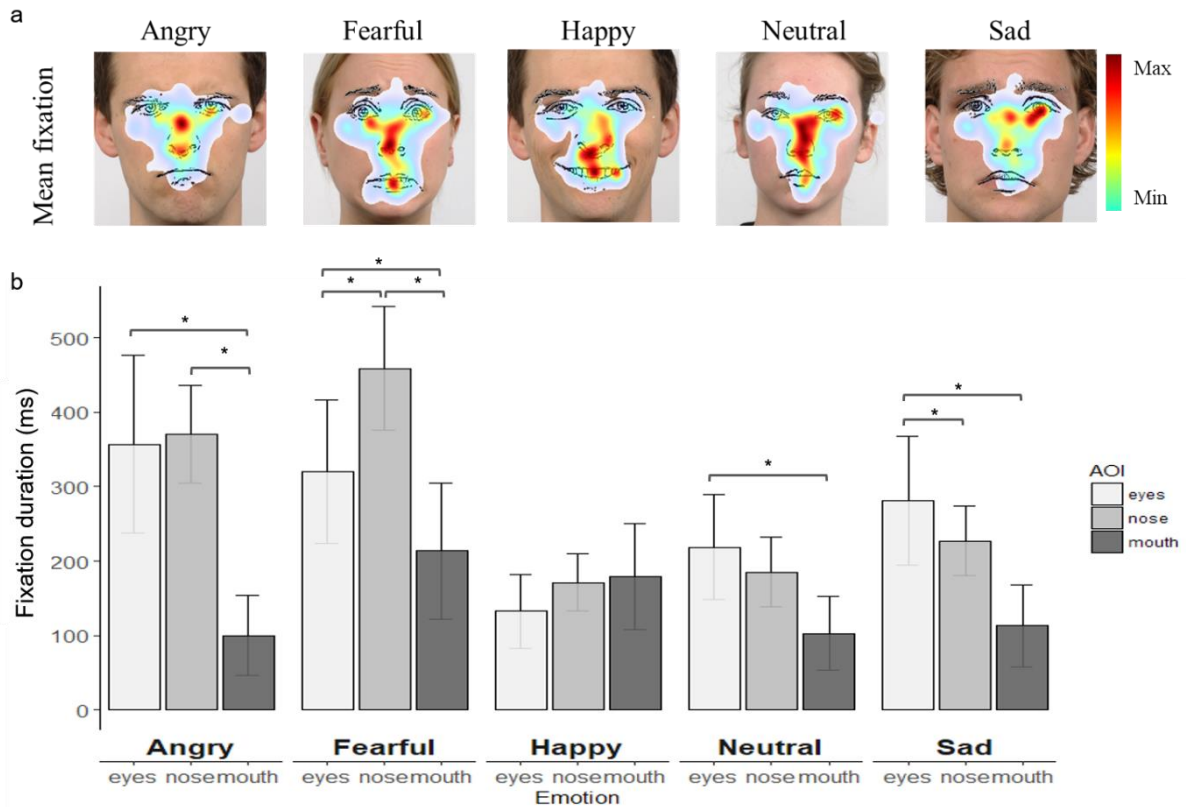


Figure 4. A) Heatmap represented the sum of fixation across participant per each emotion. The high-temperature areas (i.e., darker colour) indicate greater fixation density as opposed to low-temperature areas (i.e., lighter colour) that were occasionally visited. As there was no significant main effect of stimulus type, the images of sketch and photograph faces were combined for illustrative purposes; B) the mean fixation duration (with 95% Confidence Interval error bars) for each AOI per emotion. \*: significant ( $p < .05$ ) pairwise intra-emotion AOI comparisons.

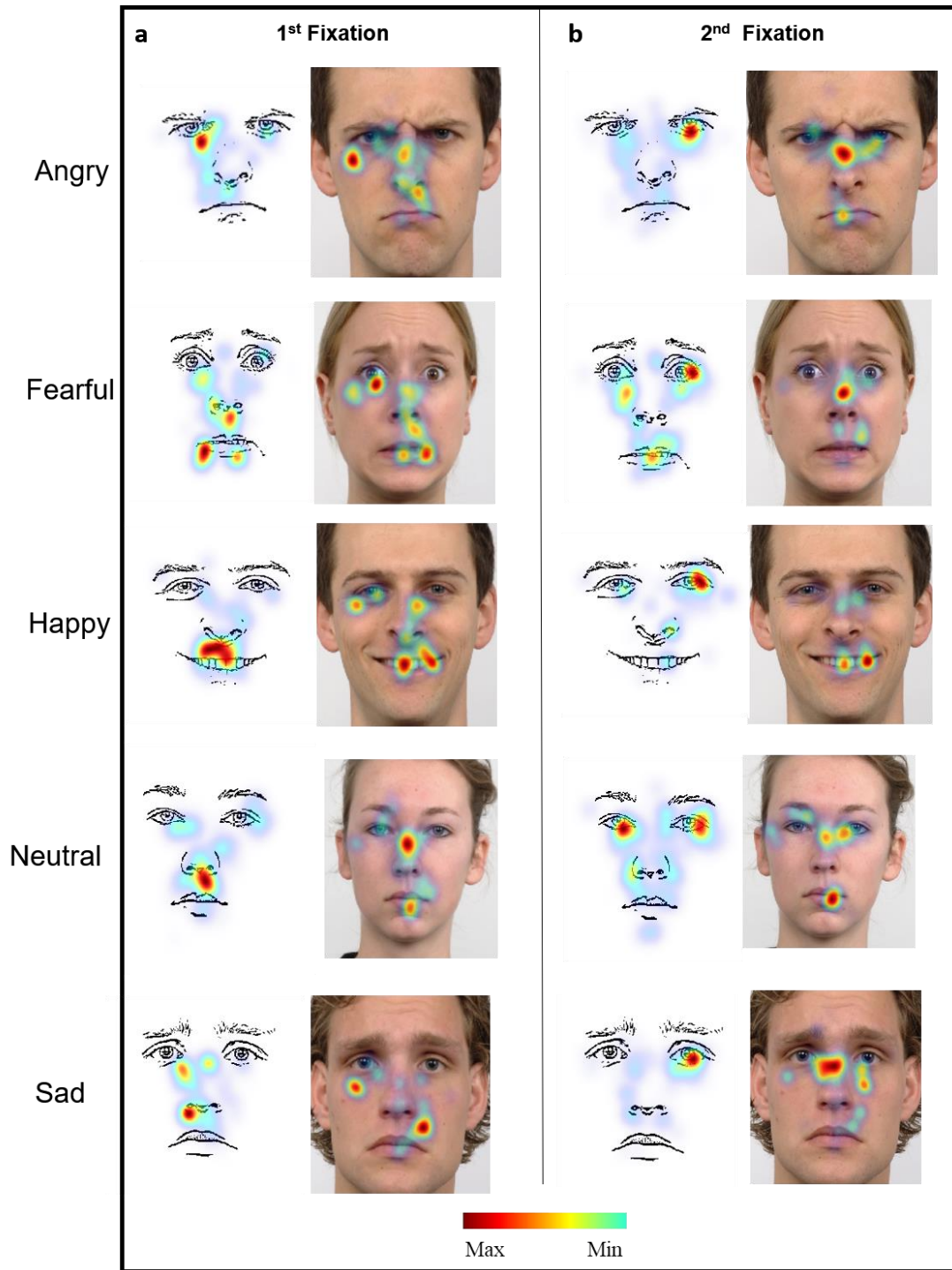


Figure 5. Heatmaps are representing the sum of fixation distribution across emotion for (a) first fixation; and (b) second fixation. Note that the first mean fixation duration and location did not vary with stimulus types across emotions. However, we provide the heatmaps of first fixation for both sketched and photographed faces, in order to compare descriptively with second fixation.

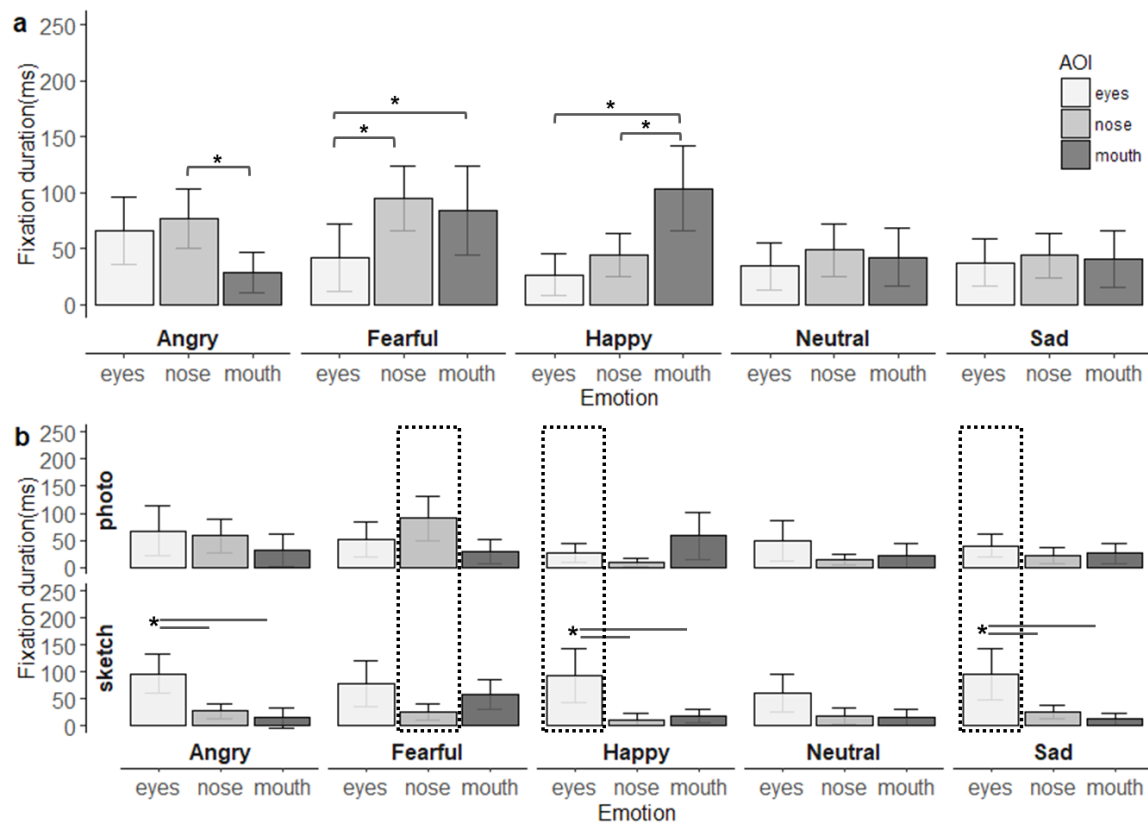


Figure 6. The mean fixation duration values for each AOI per emotion (with 95% Confidence Interval error bars) for (a) the first fixation; and (b) the second fixation. The first mean fixations across AOIs did not vary with stimulus type. However, the second mean fixations differed between stimulus type across emotions. Only the eye region on angry, happy, and sad expressions received significantly longer fixations than other AOIs across emotions on sketched faces. The dashed line demonstrated the significant difference between sketched and photographed faces. \*: significant ( $p < .05$ ) post-hoc pairwise intra-emotion AOI comparisons.

Table 1: Summary of studies analysing mean fixation duration, and first and second fixation duration in healthy subjects.

Study	Task	Emotions and AOIs	Total fixation duration	First fixation duration on an AOI	Second fixation duration on an AOI
Eisenbarth & Alpers, 2011	Valence and arousal rating, n=36. 16 expressors (KDEF) <sup>2</sup> .	Fear, anger, happiness, neutrality and sadness. Eight AOIs: forehead, left eye, right eye, left cheek, nose, right cheek, mouth, chin.	Differed between AOIs and emotions, but the left and right eye region was fixated longer than the forehead, left cheek and nose across emotions.	The location of fixation did not differ from the second fixation.	The left and right eyes were fixated more often for sadness and anger. The mouth region was fixated more for fearful, happy and neutral expressions. The second fixation focused on one of the AOIs more than the first fixation. (Rank: left eye > right eye > mouth > nose > hair > forehead > left cheek > right cheek. No significant difference between eyes and mouth. )
Vaidya et al., 2014	Emotion label and rating extreme level of facial emotion <sup>1</sup> , n=28. 12 expressors (KDEF).	Happiness, fear, disgust, surprise and neutrality. Three AOIs: eyes, nose, mouth	n/a	First fixations (measured as a proportion/frequency of total fixation duration) were located on the mouth for disgust and happiness or surprise. Other features were not targeted.	Second fixations (measured as a proportion/frequency of total fixation duration) were directed to the eye region for fear. No preferred AOI for the other emotions.
Bombari et al., 2013	Emotion recognition <sup>1</sup> , n=26. four expressors, (KDEF).	Anger, happiness, fear, and sadness. Three AOIs: eye, mouth, and nose	The nose region was fixated longer for anger than sadness. The mouth was fixated longer for happiness and anger. The longest fixation duration was found for fearful faces.	The eyes were fixated more for sadness than happiness. The mouth region was fixated more for fear than anger.	The nose was fixated more than the mouth. The eye region was fixated less for happiness than in all other emotions.
Schurgin et al., 2014	Intensity of emotional faces <sup>1</sup> , n=51. 12 expressors in grey scale, (MSFDE) <sup>3</sup> .	Happiness, disgust, fear, anger, sadness, shame and neutrality. Five AOIs: eyes, nasion region, upper nose, lower nose, upper lip.	The eyes were fixated the most across emotions, except for happiness. The nasion region was fixated relatively more for recognising fear and anger.	First fixation was on the upper nose across all emotions, however fixation on the lower nose region varied with emotion.	Second fixation on upper lip for happiness and disgust. Longer fixations on nasion region than on other AOIs.

Footnotes: 1) central fixation cross was used before stimulus onset. 2) KDEF = Karolinska Directed Emotional Faces set (Lundqvist, Flykt, & Öhman, 1998).3) MSFDE = Montreal Set of Facial Displays of Emotion (Beaupré & Hess, 2005).

## **Supplementary material: Preliminary Experiment**

The purpose was to validate firstly whether sketched faces convey sufficient information for recognizing emotions. A set of sketched facial stimuli was built from the Radboud Faces Database (RaFD) (see Langner et al. 2010) and tested for the recognition of fearful, angry, happy, sad, and neutral expressions.

### *Method*

#### *participants*

Twenty (13 male, seven female) participants with normal or corrected-to-normal vision gave their informed consent to participate in this experiment. Their mean age was 28.4 years (SD=6.7, age range 19-42). This experiment was conducted in accordance with the Declaration of Helsinki.

#### *material*

We selected frontal view photographs of 38 Caucasian adults (19 of whom were female, 19 were male) from the RaFD to equalize the gender. Both the head and the gaze faced the camera and five emotions were displayed (i.e., fear, anger, happiness, neutrality, and sadness). The creation of the emotional sketched face stimuli involved two steps: pre-processing and transformation. In the pre-processing step, a “sketch filter” was used with OpenCV Library to transform stimuli in the photographs into grayscale sketch facial stimuli. Then, the central facial region was cropped from the original RaFD 681 x 1024-pixel area to a 400 x 400-pixel area corresponding roughly to the facial contours (see Figure1a and Figure1b).

The purpose of the transformation step was to keep the areas corresponding to the diagnostic facial features, so-called AOIs: eyes (left eye & eyebrow + right eye & eyebrow), nose, and mouth (see Figure 1 in Main experiment). AOIs were manually delineated for each

stimulus. Elements such as wrinkles or black mole/spot were excluded for the AOIs. Figure 1 Suppl. illustrates the steps leading to the sketched face stimuli.

### *apparatus*

The participants were tested individually in a quiet room. The stimuli were 400 x 400 pixels subtending a 10° x 10° visual angle displayed against a 1280 x 1024-pixel white background. The stimuli were presented on a 14-inch laptop monitor with a refresh rate of 60 Hz using E-prime (Psychology Software Tools, Pittsburgh, PA). The distance between the participant's eye position to the screen was 60 cm.

### *procedure*

The experiment involved two practice sessions and one experimental session. The practice sessions were subdivided into a response key learning session (session 1) and a sketch emotion recognition session (session 2). Each practice session involved response feedback. To continue to the experimental session, the participants were required to reach 100% accuracy in session 1, then 75% accuracy in session 2. The procedure for the experimental session was the same as practice session 2 except that the sketch stimuli were different, and the participants received no feedback on response accuracy. In each session, either each practice or experimental, the trial began with a 500 ms fixation cross, followed by the stimulus, which remained on the screen until the participant responded (self-paced). This was followed by a 500 ms blank screen. The task was self-paced with no time limit. The participants were instructed to respond as quickly and accurately as possible.

Response key learning Practice session 1: This session ensured that the participants were able to correctly associate a given response key with a specific emotional state. The “D”, “F”, “H”, “J” and “K” keys of an AZERTY keyboard were used, with some keys re-labelled with stickers as follows: “C” (for “Colère”, meaning “anger” in French); “P” (for “Peur” = fear); “N” (for “Neutre” = neutral), “T” (for “Tristesse” = sadness), and “J” (for

“Joie” = happiness). This session consisted of a minimum of 30 trials with one emotional state word randomly displayed on a screen to which participants responded by pressing the corresponding key. After each five-trial block, the average response accuracy (in %) was displayed. This response key learning session continued until the participants attained 100% accuracy over five successive blocks.

Emotion recognition from sketched face, practice session 2: each trial consisted of one stimulus (either a female or male expressor) presented in the centre of the screen, randomly expressing one of five facial expressions. The participants pressed the corresponding key in response to the expression being displayed. After each block of five trials, the average response accuracy (in %) was displayed. There were six blocks in total. Participants obtaining less than an average of 75% accuracy were excluded from the rest of the study. The practice session contained ten images (2 Expressors x 5 Emotions), that had not been used in the experimental session.

Emotion recognition from sketched face, experimental session: The procedure was the same as practice session 2 except that the sketch stimuli were different, and the participants received no feedback on response accuracy. Each session included 180 randomized sketched faces (36 Expressors x 5 Emotions). There were four blocks of 45 trials separated by self-paced pauses.

#### *analyses*

Unbiased hit rates (Hu) and reaction times for correct response trials (RT) were analysed for each stimulus. Trials with RT greater than “mean + 3 SD” (9515 ms) were excluded from analysis, representing 1.8% of all trials. Furthermore, trials with RT between “mean + 2 SD” and “mean + 3 SD” were replaced by “mean + 2 SD” (7039 ms), corresponding to 1.63% of all trials (similar to Barnett, Barnett, & Lewis, 1978; Koldewyn, Jiang, Weigelt, & Kanwisher, 2013). Repeated measures ANOVAs were conducted on RT and Hu with

Emotion as a within-subject factor (fear, anger, happiness, neutrality, and sadness), followed by post-hoc Tukey (HSD) comparisons when the main effect was significant. Greenhouse-Geisser corrections were applied when the sphericity assumption was violated ( $\epsilon$  values are given in this case). Moreover,  $t$  tests were performed using the Bonferroni correction method to determine whether  $H_u$  exceeded the chance level of .20, and the criterion for the accurate recognition of  $H_u = .75$ . A detection performance level of .75 is rather conservative given that it is classically used to define a threshold stimulus level for satisfactory detection in two-alternative forced choice psychophysical paradigms (i.e., with a chance level of .50 for recognition). Here, however, the chance level was initially much lower.

### ***Results***

Two analysis steps were used on behavioural data. A first analysis was run on the 180 initial stimuli (36 Expressors x 5 Emotions), termed the sketched face database. To choose the less ambiguous sketched face stimuli for our Main experiment, we selected a subset of 50 stimuli (10 Expressors x 5 Emotions) that induced high accuracy performance (see below), out of the larger ( $n=180$ ) sketched face database. A second analysis was then performed on the selected sketched stimuli subset.

*behavioural performance for the sketched face database (180 stimuli).* The unbiased hit rate ( $H_u$ ) varied significantly with emotion,  $F(4, 76) = 136.53$ ,  $p < .0001$ ,  $\eta^2_p = .88$ . Post-hoc tests indicated that only anger and sadness did not differ for  $H_u$  ( $p = .99$ ); all other pairwise comparisons were significant ( $ps < .008$ ) (see Table 1 Suppl.). The mean  $H_u$  for each emotion was significantly above the chance level (0.20),  $ps < .01$  (using  $t(19)$  tests with a Bonferroni correction of  $\alpha = 0.05/5$ ). A satisfactory recognition rate should be greater than  $H_u = 0.75$ ; thus, we ran additional Bonferroni-corrected comparisons that showed that recognition of anger, neutrality, and sadness was significantly below this threshold (all  $ps < .01$ ), whereas recognition of fear and happiness was significantly above it (both  $ps < .01$ ). RTs also varied

significantly with emotion ( $F(4, 76) = 26.52$ ,  $\epsilon = 0.56$ ,  $p < .0001$ ,  $\eta^2_p = .58$ ). Post-hoc tests showed that the participants recognized angry faces more slowly ( $ps < .001$ ), and happy faces more quickly ( $ps < .0001$ ) than the other emotions. Neutral expressions and those of sadness and fear did not differ significantly with regard to RTs ( $ps > .93$ ) (see Table 1 Suppl.).

*behavioural performance for the sketched face subset (50 stimuli).* Although the Hu of the sketched face database was above the chance level, we considered that well-recognized sketched face stimuli should lead to Hu greater than .75 to carry emotion-specific content. Accordingly, we ranked the Hu and RT performance for each stimulus to select ten face identities per emotion as a subset of sketched face stimuli. Hu was used as the first criterion and RT was used as the second criterion. To balance the gender of face identity per emotion, the top five face identities per gender per emotion were retained. Bonferroni-corrected comparisons showed that recognition of anger, neutrality, fear, happiness and sadness was significantly above the Hu = .75 threshold (all  $ps < .01$ ). As expected, Hu for the sketched face subset was greater than the sketched face database ( $ps < .01$ ) (see Table 2 Suppl.).

A repeated-measures ANOVA on Hu revealed a significant effect of emotion ( $F(4, 76) = 11.87$ ,  $\epsilon = 0.61$ ,  $p < .0001$ ,  $\eta^2_p = .38$ ). Post-hoc comparisons showed that anger and sadness did not differ with regard to Hu ( $p > .5$ ), see Table 2 Suppl., and were similar to the sketched face database result. However, the accuracy levels for the fearful and the neutral expressions were closer to that of happiness (i.e., not significantly different, with  $p = .45$  and  $p = .49$ , respectively). A repeated measures ANOVA on RT also showed a significant main effect of emotion ( $F(4, 76) = 13.41$ ,  $\epsilon = 0.62$ ,  $p < .0001$ ,  $\eta^2_p = .41$ ). Further analyses demonstrated that there was still an advantage for happy faces (recognized more quickly than the other emotions, all  $ps < .0001$ ). Recognition of angry faces improved so that RTs were similar to those for neutral and sad expressions ( $ps > .12$ ) but longer than for fear ( $p = .0001$ ).

RTs still showed no significant difference between fear, neutrality, and sadness ( $ps > .78$ ).

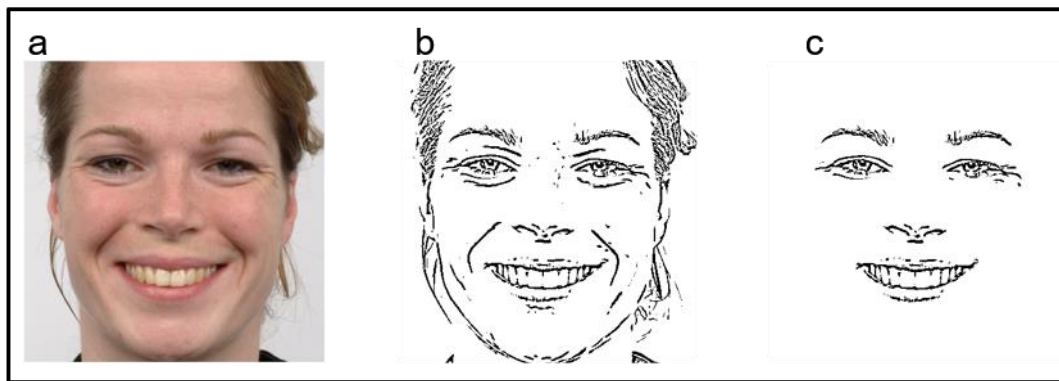


Figure 1. Suppl. From left to right: (a) a cropped 400 x 400-pixel photograph of a happy female face from the Radboud Faces Database; (b) a sketch-filtered picture; c) a sketched face with information for the eyes, nose, and mouth (e.g., AOI).

Measure	RT (ms)			<i>Hu</i>		
	<i>M</i>	95% CI		<i>M</i>	95% CI	
		<i>LL</i>	<i>UL</i>		<i>LL</i>	<i>UL</i>
Anger	2268	2126	2410	.49	.42	.55
Fear	1768	1620	1976	.85	.79	.91
Neutrality	1862	1749	1901	.59	.55	.62
Sadness	1823	1745	1901	.48	.45	.55
Happiness	1052	1018	1087	.96	.94	.97

Table 1. Suppl. Summary of RT and Hu per emotion for the sketched faces (n=180) database.

Measure	RT (ms)			<i>Hu</i>		
	<i>M</i>	95% CI		<i>M</i>	95% CI	
		<i>LL</i>	<i>UL</i>		<i>LL</i>	<i>UL</i>
Anger	2205	1727	2682	.87	.81	.92
Fear	1625	1283	1966	.96	.93	.99
Neutrality	1831	1537	2126	.96	.92	.99
Sadness	1773	1529	2017	.90	.86	.94
Happiness	931	850	1002	.99	.98	1

Table 2. Suppl: Summary of RT and Hu per emotion for the sketched faces (n=50) subset.

## 3.4 ERP responses on face inversion and deconfiguration

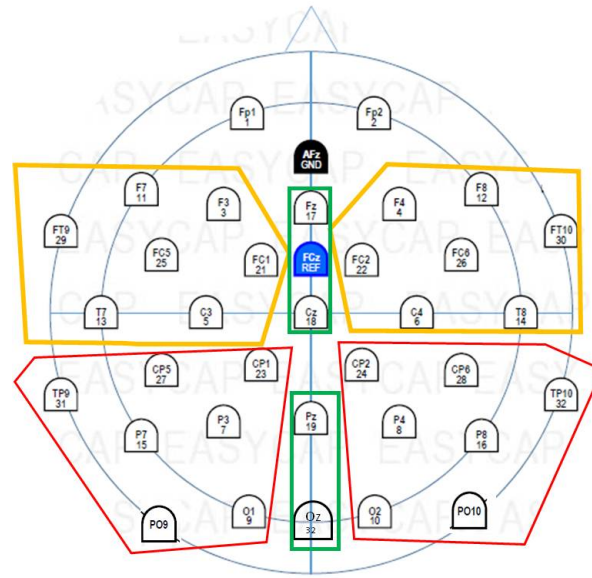
In the second half of the experimental part of this thesis, we extended the investigations to the neural responses on the created sketch face stimuli by using EEG measurements. In focus of the investigation was the manipulation of the configural information of the used stimuli by using face inversion and face deconfiguration. These two paradigms and the conducted experiments are subject to Publication 2, which is attached to the full length with supporting information in Chapter 3.5. Therefore this chapter will include only a brief summary of the hypotheses for both experiments and present a preliminary data inspection, which is not included in the Publication 2.

### 3.4.1 Preliminary EEG data inspection

The purpose of this analysis was to ensure that further ERP analyses (provided in the Publication 2, Chapter 3.5) were based on the channel that reflected optimal the visual response on face stimuli over occipital regions. This further allows a better comparison to similar studies.

#### Method

ERPs were computed exclusively on the correct answers trials. Firstly, the 31 channels were pooled into 6 regions of interest with two factors (Anterior X Posterior, Left x Middle x Right) to examine the global processing of ERPs amplitudes, see Figure 3.5. Those regions were chosen on the basis of previous experience and a face recognition study by Wiese et al. (2008). Pooling the EEG channels is an alternative way to reduce variances or averaging measurements taken from adjacent electrodes. The method of pooling is also often used to simplify ERP analysis by reducing the number of degrees of freedom or reduce problems associated with statistical testing of channels having

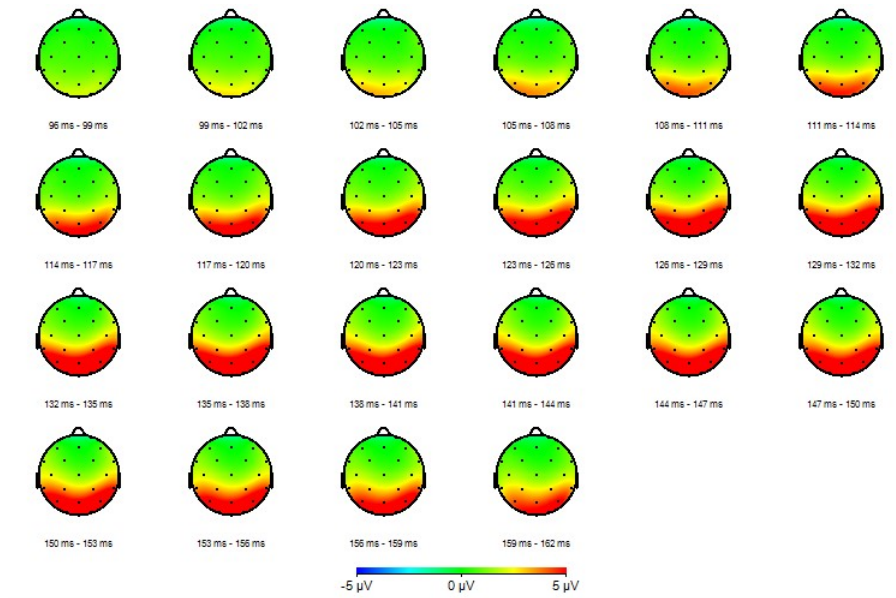


**Figure 3.5:** Schematic demonstration of pooling channels. Yellow area is anterior cortex, green area is center cortex, red area is the posterior cortex. Yellow and red regions were divided into two regions: left and right cortex. The Green region was divided into anterior and posterior cortex without reference channel.

correlated activity (Zeman et al., 2007).

## Data Analysis

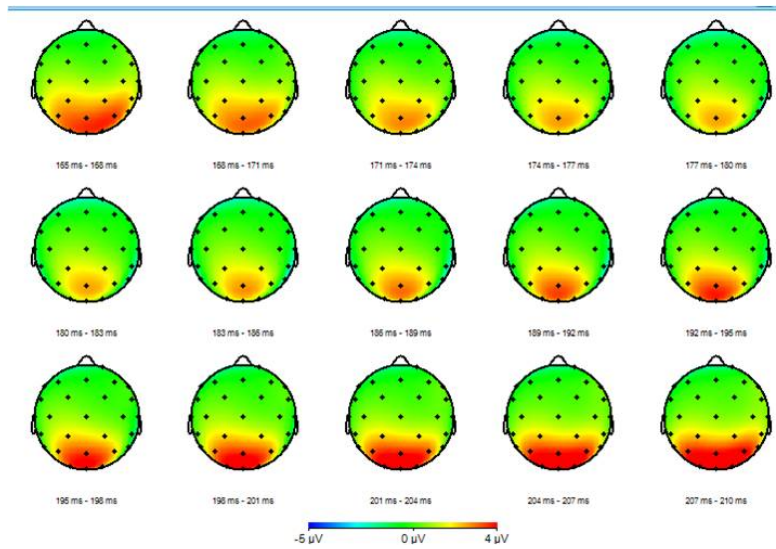
The Mixed Model ANOVA analysis were conducted for Cortical pooling group (six groups: Anterior x Posterior, Left x Middle x Right), Stimulus type (sketch and photograph faces), and Stimulus Orientation (Upright x Inverted). We would like to examine if the visual response for face stimuli over occipital regions is comparable to previous studies (e.g., Itier and Taylor, 2004b; Rossion, 2014). Therefore, we first examined especially the effect of Anterior X Posterior areas toward the activation of the P100 and N170. If the ERP values are significantly different in this case, then the interaction of Hemisphere (Left x Center x Right) X Stimulus type x Stimulus Orientation is reported. Greenhouse-Geisser corrections adjusted degrees of freedom were used when necessary ( $\epsilon$  values are given in this case). We made post-hoc Tukey's comparisons (HSD) when the main effect was significant. Data were analyzed using Statistica (StatSoft, Inc.) This analysis was used to examine the overall electrophysiological responses to sketch and photographed faces and the face inversion effect. As



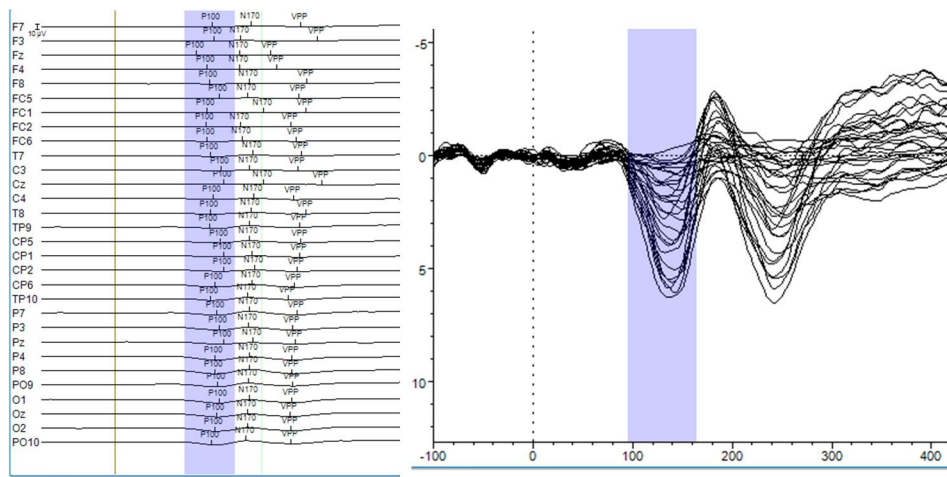
**Figure 3.6:** Topographical plots for the P100 (96ms-162ms, in three millisecond time frame) shows the posterior cortex was more active than the anterior cortex across electrodes and trials.

the main focus is to examine the Pooled channels (Anterior x Posterior) x Stimulus type x Stimulus orientation and Hemisphere (Left x Center xRight) x Stimulus type x Stimulus orientation, therefore only these two interactions will be report here. Once we confirmed the visual processing towards face stimuli occurred across the occipital area, we conducted the following ERP analysis on P100 (see Figure 3.6) and N170 components (see Figure 3.7).

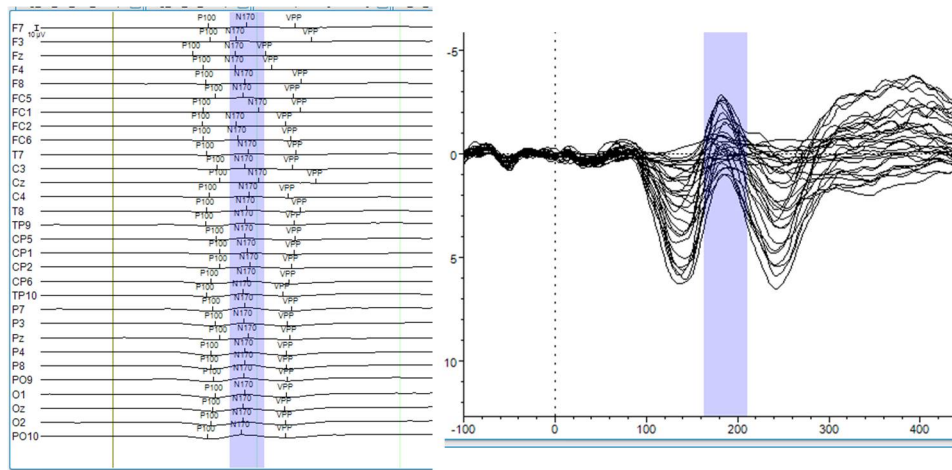
Segments were averaged separately for each stimulus type (sketch face, photographed face) and orientation (upright, inverted). We chose the single maximal value electrode from left, and right hemisphere by computing mean amplitude across subjects for P100 and N170. The voltage and latency of the P100 ERP component (O1 and O2 electrodes, 100-162 ms, see Figure 3.8), N170 (P7 and P8 electrodes, 163-210 ms, see Figure 3.9) were analyzed separately using repeated measures ANOVAs shown in Publication 2.



**Figure 3.7:** Topographical plots for the N170 (155ms-210ms, in three millisecond time frame) shows that the posterior cortex was more active than the anterior cortex across electrodes and trials.



**Figure 3.8:** Corresponding ERP traces for all electrodes as grand average across all trials. The left panel shows all P100 peaks located in the time-window of 100ms - 162ms; Right panel: time window shaded in purple covering the proximate area of each P100 ERP across all electrodes.



**Figure 3.9:** Corresponding ERP traces for all electrodes as average across all trials. Left panel shows all peaks located in the time-window of 163ms - 210ms; Right panel shows: time window shaded in purple covering the proximate area of each ERP across all electrodes.

### Preliminary EEG data results

**P100 component** The Mixed Model ANOVAs showed no significant interactions of the Pooled channels (Anterior x Posterior) x Stimulus type x Stimulus orientation  $F(1,24) < 1$ , n.s. as well as Hemisphere (Left x Center X Right) x Stimulus type x Stimulus orientation  $F(2, 48) = 3.67$ ,  $p > .56$ ,  $\eta^2_p = .11$ .

**N170 component** The Mixed Model ANOVAs on N170s showed significant interaction of Pooled channels (Anterior x Posterior) x Stimulus type x Stimulus orientation  $F(1,24) = 4.78$ ,  $p < .04$ ,  $\eta^2_p = .17$ . But this interaction did not varied with hemisphere (Left x Center x Right),  $F(2,48) = 2.02$ ,  $p > .12$ ,  $\eta^2_p = .07$ . Post-hoc Tukey analyses on the former significant interaction showed that there was no significant difference for Stimulus type and Stimulus orientation in anterior area ( $ps > .82$ ), but in the posterior cortex ( $ps < .04$ ) except there was no significant difference for inverted sketch and upright sketch faces ( $p > .98$ ). We observed only one interaction between anterior and posterior cortex, which was the significant lower amplitude for upright photograph face ( $-.078 \mu V$ ) in the anterior cortex than upright photograph faces in the posterior cortex ( $2.02 \mu V$ ).

## Discussion

The purpose of the preliminary EEG data inspection is to examine the interaction of the neural activity for each region of interest (Anterior x Posterior, Left x Center x Right) between the new sketch face stimuli and their photograph counterparts in different orientations. The results of N170s showed that sketch and photographed faces were significant larger in the occipital cortex than in anterior cortex, espeically for uprigh photographed faces. The electrophysiological responses to faces in occipital areas were supported by numerous studies [Eimer \(2000a\)](#); [Sadeh et al. \(2010\)](#). Selected channels for pooling are generally based on commonly accepted practice, *ad hoc* observation, thus, often leads to a blurring of the ERP amplitude and attenuates statistical measures between conditions ([Dien, 1998](#)). But, the pooling results allow us to examine of which region of cortex is activated to corresponding ERP component. Thus, we used the results of pooling to examine whether ERPs components were varied significantly with the stimulus type and the orientation.

### 3.4.2 Hypothesis Face inversion

From the results of the previous eye-tracking study, employment of different visual processes occurred at the early stage of face processing (the second fixation) between sketch and photographed faces, suggesting sketch faces might engage in different face processes as compared with photographed faces. The face inversion and the removal of facial texture information have been suggested to influence face processing at both behavioral and neural levels. It is still uncertain whether these specific characteristics of facial texture information (sketch and photographed faces) and face inversion interact with each other during the specific condition of emotion recognition. Therefore, in this study, we investigate the interaction between facial texture information (sketch and photographed faces) and inverted face in emotion recognition on behavioral and event-related potential (ERP) responses, notably the P100 and N170.

Given that sketch faces share the common diagnostic features (i.e., the eyes, nose, mouth) as in photographed faces, we expected the behavioral performance would be comparable. For the electrophysiological responses, we hypothesized there would be an interaction between face inversion and face stimulus type (sketch or photographed faces). Therefore, we expected that a significant N170 inversion effect would be found in the photographs but not in the sketch condition. Accordingly, measures of the latency of N170 will ascertain that upright photograph faces are associated with the most efficient processing.

### 3.4.3 Hypothesis Face deconfiguration

Based on the results obtained in the experiment of [3.4.2](#) where upright sketch elicits a larger N170 amplitude faces relative to the inverted sketch faces, a dominance of part-based processing was suggested. Additionally, the enhancement of the N170 amplitude suggested that the N170 not be merely triggered by the physical characteristic of the sketch stimuli (e.g., features), but also depends on the configural information to

distinguish the emotions. To better understand the role of diagnostic features in the contribution of emotion recognition, the P100 and N170 amplitude and latency are measured while configural information of sketch faces vary parametrically. This parametric manipulation will make it possible to determine whether the N170 component somehow reflects the cognitive load related to the processing of configural information in cases where sketch stimuli encourage part-based processing. This manipulation of facial features has not yet been demonstrated.

### 3.5 *Publication 2*

Partial perceptual insensitivity to configural manipulations of emotional faces: An ERP study

Yu-Fang Yang<sup>a,b\*</sup>, Eric Brunet-Gouet<sup>c,d</sup>, Mariana Burca<sup>c,d</sup>, Emmanuel Kalunga<sup>e</sup>, &  
Michel-Ange Amorim<sup>a,b</sup>

<sup>a</sup>CIAMS, Univ. Paris-Sud, Université Paris-Saclay, France; <sup>b</sup>CIAMS, Université d'Orléans,  
France; <sup>c</sup>HANDIRESP, UVSQ, Université Paris-Saclay, France; <sup>d</sup>Centre Hospitalier de  
Versailles, France; <sup>e</sup>LISV, UVSQ, Université Paris-Saclay

Corresponding author: [yu-fang.yang@u-psud.fr](mailto:yu-fang.yang@u-psud.fr)

Correspondence address: CIAMS, UFR STAPS, Univ. Paris-Sud, Université Paris-Saclay,  
Rue Pierre de Coubertin, 91405 Orsay Cedex, France

Affiliation's address: HANDIRESP, UVSQ, Université Paris-Saclay, 50 Rue Berthier, 78000  
Versailles, France

Affiliation's address: Centre Hospitalier de Versailles, Hôpital Mignot, 177 Rue de  
Versailles, 78157 Le Chesnay, France

Affiliation's address: LISV, UVSQ, Université Paris-Saclay, 10-12 avenue de l'Europe  
78140 Velizy, France

Funding sources: This work was supported by a PhD grant from Taiwan and Université Paris-  
Sud, and by a public grant overseen by the French National Research Agency (ANR) as part  
of the “Investissement d'Avenir” program, through the “iCODE Institute” project funded by  
IDEX Paris-Saclay, ANR-11-IDEX-0003-03.

## Insensitivity to configural manipulations of emotional faces

### Partial perceptual insensitivity to configural manipulations of emotional faces: An ERP study

#### Abstract

Removal of texture information (e.g., sketches) reduces configural processing of faces, presumably due to the brain being tuned to textured stimulation (e.g., natural or photographed faces). However, its neurocognitive consequences on processing expressive information transmitted by diagnostic features (eyes, nose, mouth) are unknown. Here, we investigated the behavioral and neuroelectric (P100 and N170) effect of configural information manipulation on emotion recognition by examining a) the inversion effect in both sketched and photographed faces (Expt 1); and b) the effect of spatial deconfiguration in upright sketched faces (Expt 2). Expt 1 showed that although upright sketched and photographed faces induced comparable high behavioral performance, the inversion effect degraded emotion recognition significantly more for inverted sketched than photographed faces. Stimulus impoverishment and inversion increased N170 peak latency. In contrast, although stimulus impoverishment increased N170 amplitudes for upright stimuli, the effect was reversed when inverted. Expt2 showed that to keep recognition accuracy stable, the processing cost increased a) linearly with the degree of sketched face deconfiguration as measured by reaction times, and b) nonlinearly on P100 latency and N170 amplitude (sudden change after a given degree of deconfiguration). Our study shows that relevant diagnostic features can convey sufficient emotional information even with altered configural information (stimulus impoverishment, or spatial deconfiguration) or processing (stimulus inversion). However, to maintain fair recognition performance, the processing cost will be increased at both the behavioral and neuroelectric levels, suggesting that emotion recognition depends on both configural and part-based processing.

*Keywords:* N170; P100; configural processing; part-based processing; emotional facial expression

## 1. Introduction

Visual recognition of facial expressions is a crucial element of social communication. The ability to recognize emotion successfully requires in-depth processing of facial diagnostic features such as the eyes, nose, and mouth to distinguish one emotion from another (Adolphs, 2002; Smith, Cottrell, Gosselin, & Schyns, 2005). The ability to recognize facial expressions is accomplished by processing configural information and part-based information (Meaux & Vuilleumier, 2016; Tanaka, Kaiser, Butler, & Le Grand, 2012). Part-based processing refers to the extraction of partial information successively from local features. Configural and holistic processing is thought to integrate facial features simultaneously, although the distinction is under debate (Piepers & Robbins, 2012). Configural processing takes into account the spatial distances and relative positioning of local facial features, whereas holistic processing focuses on the integration of several features into a “gestalt” representation (Goffaux & Rossion, 2006; Maurer, Le Grand, & Mondloch, 2002; McKone & Yovel, 2009). Recent eye-tracking (Bombari et al., 2013), and fMRI studies (Meaux & Vuilleumier, 2016) have shown that all these processes contribute to emotion recognition. To better understand emotional face processing, the literature on the perception of non-emotional faces is particularly relevant. In the introduction, we will provide a selective review of paradigms that modify the balance between configural and part-based processing, before considering electrophysiological evidence supporting the neural validity of these experimental manipulations. Last, some frameworks stemming from cognitive neuroscience will be briefly presented to provide a heuristic information-processing model within which our findings fit.

First, a consensual method for impairing configural processing consists in presenting facial stimuli upside down, known as face inversion (Namdar, Avidan, & Ganel, 2015; Yin,

## Insensitivity to configural manipulations of emotional faces

1969). This unusual transformation of the coordinate frame paired with the configural variables leads to decreased facial recognition performance (Bartlett & Searcy, 1993; Helmut Leder & Bruce, 2000), and favors part-based processing at the expense of configural processing (Le Grand, Mondloch, Maurer, & Brent, 2001; Richler, Mack, Palmeri, & Gauthier, 2011). On the other hand, low-level visual information also has an impact at the behavioral level and might change differentially part-based and configural processing. For instance, photographed faces are identified more accurately than sketched ones in a forced-choice face identification task, because sketches reduce depth cues provided by facial texture and shadows (Leder, 1999; Tzschaschel, Persike, & Meinhardt-Injac, 2014). The removal of the facial surface (e.g., shadows and color) is generally considered to evoke preferentially part-based processing (Eng et al., 2017; Zhao, Bülthoff, & Bülthoff, 2016). The facial texture and shape of natural faces provide crucial surface information for features binding into a configuration smoothly (Meinhardt-Injac, Persike, & Meinhardt, 2013). Several studies have found that participants performed better in a face identification task using real faces (photograph and grayscale) rather than sketched faces (line-drawing) (Tzschaschel et al., 2014). However, some behavioral studies demonstrated that participants identified sketched faces more quickly and accurately than photographed natural faces when the task required identifying differences in local facial features, such as the composite face task (Zhao et al., 2016) or the feature change detection task (Meinhardt-Injac et al., 2013)

In addition to behavioral performance, the inversion and modification of facial texture also influence neural mechanisms correlated to face processing. The amplitude and peak latency of event-related potentials (ERP) have proven useful for investigating the underlying neural mechanism of face processing. The earliest well-replicated component, the P100, peaks at around 100 ms (Sadeh & Yovel, 2010) and is sensitive to low-level visual parameters such as luminance and color (Regan, 1989). The P100 has been associated with

## Insensitivity to configural manipulations of emotional faces

configural processing (Wang & Fu, 2018; Wang, Sun, Ip, Zhao, & Fu, 2015). The greater, and delayed, P100 component is triggered by inverted rather than upright faces (Batty & Taylor, 2003; Feng, Martinez, Pitts, Luo, & Hillyard, 2012; Herrmann, Ehlis, Ellgring, & Fallgatter, 2005; Itier, Latinus, & Taylor, 2006). Following the P100, the N170 is a face-sensitive negative occipitotemporal deflection that peaks at around 170 ms post-stimulus onset (Bentin, Allison, Puce, Perez, & McCarthy, 1996; B Rossion et al., 2000; Bruno Rossion et al., 1999), independently of the task or the familiarity of the faces (Anaki, Zion-Golumbic, & Bentin, 2007). The N170 is significantly affected by facial orientation, that is, the N170 is greater and delayed during processing of an inverted face in comparison to an upright one (Bentin, Golland, Flevaris, Robertson, & Moscovitch, 2006; Caharel, Collet, & Rossion, 2015; Itier et al., 2006; Minami, Nakajima, Changvisommid, & Nakauchi, 2015). Interestingly, some studies reported that the N170 face inversion effect is absent for sketched faces in comparison with photographed faces (Sagiv & Bentin, 2001), high-spatial frequency facial stimuli (Flevaris, Robertson, & Bentin, 2008; Goffaux, Jemel, Jacques, Rossion, & Schyns, 2003), and isolated facial features or the eyes (Itier, Alain, Sedore, & McIntosh, 2007; Sagiv & Bentin, 2001). Taken together, these results suggest that the reduction of visual information from photographed faces to sketched faces could diminish configural information, as could face inversion that would disrupt early face processing.

Using fMRI, a recent investigation offered new evidence showing that facial expression recognition might engage separate brain networks depending on the processing mode (Meaux & Vuilleumier, 2016). Using composited emotional faces, the authors showed that configural processing integrating expression information activated the fusiform, the inferior occipital areas and the amygdala when internal features were congruent (i.e., face template matching). Conversely, the analysis of local part-based information was associated with superior temporal sulcus (STS) and prefrontal area activations in the context of full-face

## Insensitivity to configural manipulations of emotional faces

configurations. Although the precise neural substrates of the P100 and N170 are under debate, one study has shown that the former component is at least partially generated within the fusiform gyrus and reflects first “face template matching”, whereas the latter engages the gyrus fusiformis and represents face processing in the occipital cortex (Herrmann et al., 2005). These findings are supported by a number of hierarchical models of face perception that associate the earliest components with the activities of the occipital face areas, the fusiform face areas and the superior temporal sulcus (Pitcher, Walsh, & Duchaine, 2011).

Returning to the question raised here on the neurocognitive mechanisms for processing emotional faces, it is useful to point out that the paradigms are based on the recognition of a limited number of categories (here, basic emotions) that are invariant with respect to the identity of the individuals. In these terms, simple decision-making models can be used to link behavioral variables (emotional recognition performance) and the processing of the information presumed to be involved. Since proposing an integrative localization model would fall outside the scope of this work, we will limit our account to an approach that links the presumed visual information processes and the P100 and N170 electrophysiological variables. Our argument is that during emotion recognition, representations (either stored face templates or featural information) of given emotions compete for the final decision (Palmeri & Cottrell, 2010; Palmeri, Schall, & Logan, 2015). Accordingly, the closer the visual input from the quality of stimulation/information to which the visual system is tuned, the easier/faster the visual processing and the decision process will be. In a simplification approach, we assume that a decision model (i.e., an accumulator model) reflecting differences in response times can be related to the balance between the considered P100 and N170 components, in terms of both amplitudes and latencies. The electrophysiological components are thought to reflect configural/holistic processing of emotional faces and diagnostic

Insensitivity to configural manipulations of emotional faces

features, respectively, through specialized template recognition systems. The model will be described in light of our experimental results in the discussion.

### 1.1. Objectives

The evidence mentioned above suggests that face inversion and the removal of facial texture influence face processing at both the behavioral and neural levels. It is still uncertain whether these specific characteristics of facial texture information (photographed and sketched faces) and inverted face interact during emotion recognition. In addition, experimental evidence ranging from behavior to electrophysiological measures are difficult to account for within an integrated neurocognitive model. The present study will be discussed in an attempt to model how emotion recognition is processed at both the global/configural and featural levels with a balance based on stimulus types, orientation, and configural integrity. To do so, it was necessary to design a novel sketch stimuli set that retained only diagnostic features without facial texture or noninformative parts, adapted from the Radboud Face Database (RaDF) (Langner et al., 2010), making it possible to match photographed face counterparts. Since the sketched face stimuli (without head contour) share common diagnostic features (i.e., the eyes, nose, mouth) with their photographed counterparts, we made the preliminary assumption that behavioral performance would be equivalent at least in canonical upright orientations.

At the electrophysiological level, in the first experiment, we hypothesized an interaction between the effect of face inversion and visual information reduction (from photographed to sketched faces), based on the results mentioned earlier (Caharel et al., 2013; Sagiv & Bentin, 2001). Concerning P100, we expected that the P100 would reflect low-level processing and be sensitive to stimulus type more than to orientation. We assumed that, because sketched facial stimuli do not elicit efficient configural processing (due to

Insensitivity to configural manipulations of emotional faces

impoverished information) in early face processing, these stimuli would recruit additional neural resources for part-based processing in comparison with upright photographed faces. Therefore, we expected the N170 modulation to provide a relevant biomarker for this deeper processing, in terms of peak latency, amplitude or both. To further explore the role of configural processing for impoverished emotional facial stimuli, in a second experiment we investigated ERP components evoked by sketched faces while configuration information (spatial relationships between diagnostic features) was parametrically altered (this procedure will be referred to as “deconfiguration” henceforth). We expected a corresponding parametric modulation of reaction times and aimed to determine whether the P100 and N170 components are affected by this parameter.

## 2. Experiment 1

### 2.1. Materials and methods

#### 2.1.1. *Participants*

Twenty-five healthy volunteers (six females and 19 males, mean age:  $26.4 \pm 6.5$  years, age range: 20-40 years) with normal or corrected-to-normal-vision were recruited from the Université Paris-Sud. Twenty subjects were right-handed according to a translated version of the Edinburgh Handedness Inventory (Oldfield, 1971). All of the subjects gave their informed consent before participating. The study was conducted in accordance with the Declaration of Helsinki.

#### 2.1.2. *Stimuli presentation*

Two sets of facial expression stimuli were used in the experiment: sketched facial stimuli and their photographed face counterparts chosen from RaFD (Langner et al., 2010). We applied a sketch filter in OpenCV to the photographed facial stimuli to create a set of sketched stimuli retaining the eyes, nose, and mouth diagnostic features, and excluding

### Insensitivity to configural manipulations of emotional faces

insignificant features such as hair, spots, etc. These two subsets were validated and selected from previous behavioral and eye-tracking experiments on emotion recognition using reaction times and unbiased hit rate as criteria and fixation measures (Yang et al., submitted). Each stimulus set contained five facial expressions: fearful, angry, sad, happy, and neutral. Each emotion category included five upright and five inverted facial stimuli, and identity gender was randomized and counterbalanced across emotions. All of the images were enclosed within a rectangular frame, in a 400 x 400 pixel array. Facial stimuli subtended a 10° x 10° visual angle and were presented centrally on a 17-inch computer monitor (resolution 1280 x 1024 pixels) with a 60 Hz refresh rate. The value of the visual angle conformed to that seen in normal face-to-face interaction (Henderson, Williams, & Falk, 2005). We used E-prime 2.0 software to conduct the experiment.

#### 2.1.3. Procedure

The task was a forced-choice facial emotion recognition task whereby participants were required to select the emotion displayed by a facial expression stimulus from among five possible responses: neutrality, anger, fear, happiness and sadness. The experiment consisted of two successive parts: (1) practice session (see Supplementary material for details) and (2) experimental session.

The 'Experimental session' comprised two blocks (one for sketched faces and the other for photographed faces) of 150 trials each. The procedure was similar to the practice session with the participant pressing the key corresponding to the displayed emotion, except that no feedback on performance was provided. There were 50 stimuli per block: ten10 expressors (5 upright, 5 inverted) x 5 emotions, and each was repeated three times randomly within blocks. Each block was separated by a pause, and block order was counterbalanced across participants. Each trial began with a 600 ms fixation cross, then a 300 ms blank screen (used as a baseline period for EEG analysis), followed by the facial stimulus displayed

Insensitivity to configural manipulations of emotional faces

against a white background, which remained on the screen until the participant responded. Although the response time was self-paced with no time limit, the participants were instructed to respond as quickly and accurately as possible, and to avoid blinking during the displayed stimulus.

#### *2.1.4. EEG acquisition*

EEG was recorded using BrainAmp amplifier and Brain Recorder software (Brain Products, Munich, Germany) with 32 Ag/AgCl active scalp electrodes (Acticap). The electrodes were referenced to a common ground situated at the vertex. Eye movement was monitored by VEOG (vertical electrooculogram) with FP1 and FP2. Electrode impedances were maintained below 10k. The EEG signals were measured at a sampling rate of 1000 Hz and were filtered through a 50-Hz notch filter. Data were bandpass filtered offline between 0.1 and 40 Hz (24 dB per octave). To eliminate electrical artifacts, epochs with voltages exceeding  $\pm 50 \mu\text{V}$  at any electrode were excluded from analysis. Ocular artifacts were removed using ocular correction ICA after examining data for artifacts and marker respective time periods using raw data inspection. An average of 63 trials were retained per condition (range 53-67) across 25 participants. Epochs were taken from 100 ms before visual stimuli onset to 1000 ms afterward. The baseline was defined as the average voltage during the [-100 ms; 0 ms] post-onset stimulus time window.

#### *2.1.5. Data analysis*

Regarding behavioral data, first, accuracy was calculated as an unbiased hit rate (Hu) for each emotional state category (Wagner, 1993) to control for potential response biases (Goeleven, De Raedt, Leyman, & Verschuere, 2008). Second, only correctly answered trials were used for the analysis of reaction times (RTs), and outliers as the values greater than mean +3SD (i.e.,  $> 3357 \text{ ms}$ ) were excluded (5.3% of total trials). Furthermore, 9.8% of the trials that contained  $\text{mean} + 2\text{SD} \leq \text{RT} \leq \text{mean} + 3\text{SD}$  were replaced with mean+2SD

## Insensitivity to configural manipulations of emotional faces

(2238 ms) from our analysis in order to reduce the influence of outliers. Repeated-measures ANOVAs were performed on RT and Hu with stimulus type (sketch, photograph) and orientation (upright, inverted) as within-subject factors, and followed by post-hoc Tukey (HSD) comparisons when effects were significant. We also provide 95% Confidence Intervals and Cohen's  $d$  values as measures of effect size.

ERPs were computed exclusively on the correctly answered trials with less than mean RT +3SD in order to compare behavioral results. The segments were averaged separately for each stimulus type (sketch, photograph) and orientation (upright, inverted). The voltage and peak latency of the P100 ERP component (O1 and O2 electrodes, 100-162 ms), and N170 (P7 and P8 electrodes, 163-210 ms) were analyzed separately using repeated measures ANOVAs. For this analysis, the within-subject factors were stimulus type (sketched face, photographed face), stimulus orientation (upright and inverted) and hemisphere (left, right). Post-hoc Tukey's comparisons (HSD) were used for significant effects involving more than two means. Data were analyzed using Brain Analyzer (Brain Products, Munich, Germany) and Statistica software (StatSoft, Inc.). We set the reference significance threshold at 0.05 for behavioral and EEG data analyses.

## 2.2. Results

### 2.2.1. Behavioral data

ANOVA revealed significantly greater accuracy (Hu) for photographed facial stimuli ( $M = .86$ ) than for sketched faces ( $M = .75$ ),  $F(1,24) = 131.32$ ,  $p < .001$ ,  $\eta^2_p = .85$ . Hu deteriorated with stimulus inversion,  $F(1,24) = 123.84$ ,  $p < .001$ ,  $\eta^2_p = .84$ , with  $M = .90$  for upright stimuli and  $M = .71$  for inverted faces. The Stimulus type x Orientation interaction was significant  $F(1,24) = 67.77$ ,  $p < .001$ ,  $\eta^2_p = .74$ . Post-hoc analysis indicated that accuracy for upright photographed ( $M = .91$ ) and upright sketched faces ( $M = .89$ ) did not differ significantly at  $p = .42$ . The stimulus inversion effect was found for both the photographed

### Insensitivity to configural manipulations of emotional faces

(inverted face  $M = .81$ ) and sketched faces (inverted sketch  $M = .61$ ),  $ps < .001$ , although more for sketched faces ( $M_{\text{effect}} = .26$ ) than photographs ( $M_{\text{effect}} = .10$ ), see Figure 1 A.

ANOVA on RTs revealed no mean difference between photographed ( $M = 1260$  ms) and sketched stimuli ( $M = 1271$  ms),  $F(1,24) = 0.2$ ,  $p = .66$ ,  $\eta^2_p = .008$ . However, stimulus inversion significantly increased RTs,  $F(1,24) = 35.8$ ,  $p < .001$ ,  $\eta^2_p = .60$ , with  $M = 1220$  ms for upright stimuli and  $M = 1311$  ms for inverted faces. The Stimulus type x Orientation interaction was significant,  $F(1,24) = 5.96$ ,  $p = .02$ ,  $\eta^2_p = .20$ . Post-hoc analysis indicated that stimulus inversion increased RTs significantly for photographed ( $p < .04$ ), and sketched faces ( $p < .001$ ). However, there was no stimulus type effect for either upright ( $p = .66$ ) or inverted faces ( $p = .13$ ), see Figure 1B.

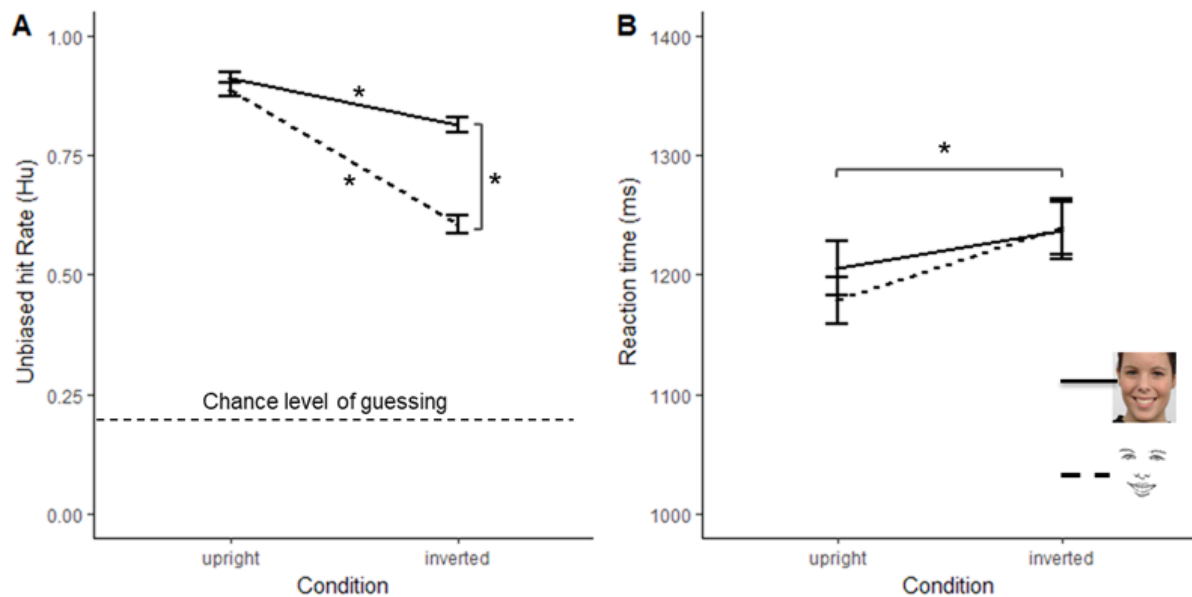


Figure 1. Stimulus type x Orientation interaction on (A) average Hu with chance level of .20; and on (B) reaction times. \* indicates  $p$  value  $< .05$ . Error bars indicate standard error of the mean.

### 2.2.2. ERP result

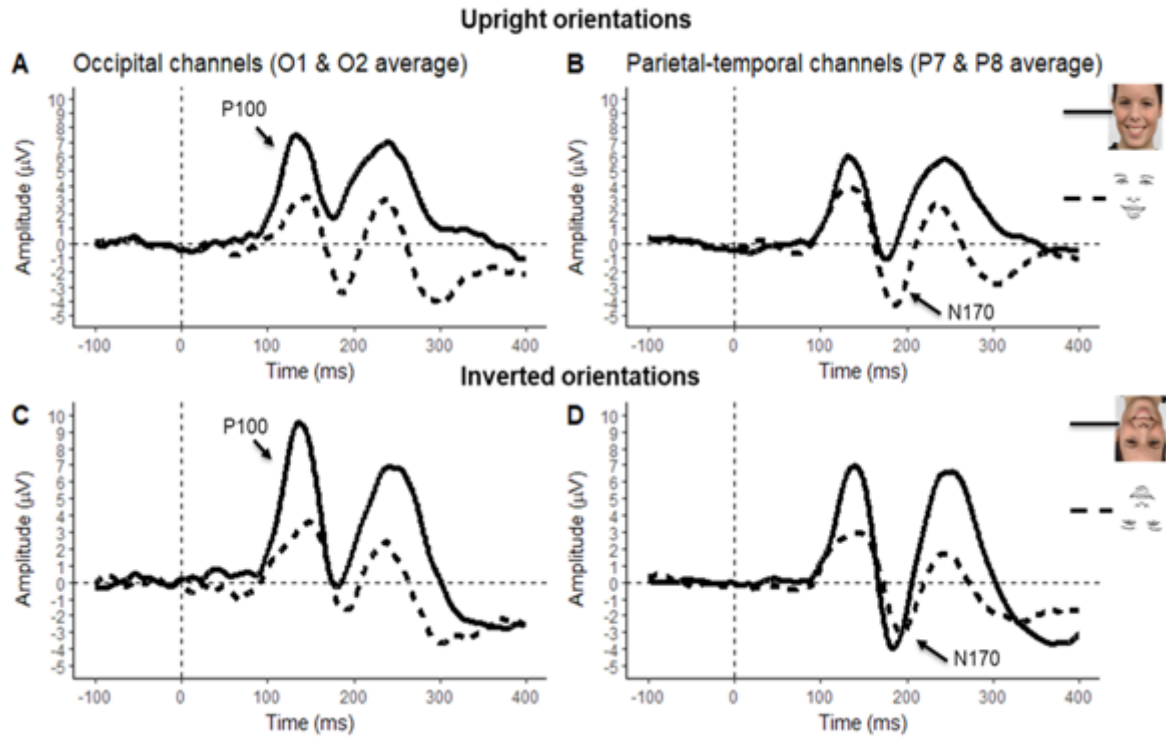


Figure 2. Grand average ERP waveforms elicited by photographed and sketched faces in the upright (top) and inverted (bottom) orientations at the averaged left and right occipito-temporal electrode site (O1 & O2) for the P100 component (A & C) and the left and right parietal-temporal electrode site (P7 & P8) for the N170 component (B & D).

#### 2.2.2.1. P100 amplitude

The ANOVA showed greater P100 amplitude in response to photographed faces ( $M = 7\mu\text{V}$ ) compared to sketched faces ( $M = 3\mu\text{V}$ ),  $F(1,24) = 79.27$ ,  $p < .001$ ,  $\eta^2_p = .77$ , and in response to inverted stimuli ( $M = 6\mu\text{V}$ ) compared to upright stimuli ( $M = 5\mu\text{V}$ ),  $F(1,24) = 11.1$ ,  $p < .003$ ,  $\eta^2_p = .32$  (see Figure 2 A&C). There was a nonsignificant Stimulus type x Orientation interaction,  $F(1,24) < 1$ , n.s., with a parallelism pattern suggestive of an additive effect of Stimulus type ( $4\mu\text{V}$  increase for photographed faces) and Orientation ( $1\mu\text{V}$  increase for inverted facial stimuli) on P100 amplitude, see Figure 3. It is interesting to note that the size of stimulus type effect (95% CI =  $[3.2\mu\text{V}; 5.1\mu\text{V}]$ ,  $d=1.76$ ) was much greater and robust

Insensitivity to configural manipulations of emotional faces

than the moderate inversion effect (95% CI = [0.4 $\mu$ V; 1.7 $\mu$ V], Cohen's  $d = 0.66$ ). Finally, there was no effect of Hemisphere on P100 amplitude, nor any other significant interaction between factors, all  $ps > .20$ .

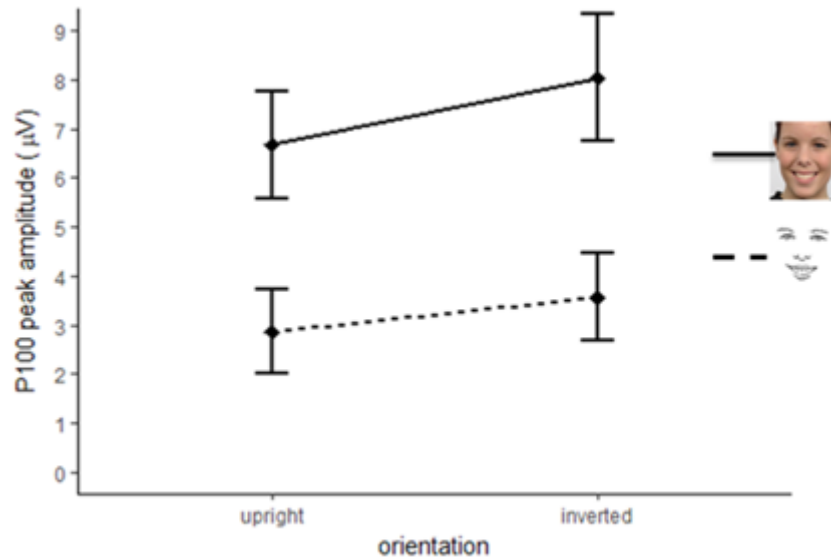


Figure 3. Mean P100 peak amplitude for each Stimulus type x Orientation. Error bars indicate standard error of the mean.

#### 2.2.2.2. P100 peak latency

The ANOVA showed no main effects,  $ps > .22$ . There was no interaction between Hemisphere and Stimulus type, nor between Stimulus type and Orientation,  $ps > .08$ . However, we found a significant Hemisphere x Orientation interaction,  $F(1,24) = 7.94$ ,  $p = .01$ ,  $\eta^2_p = .25$ . Post-hoc comparisons showed delayed P100 peak latency only in the left hemisphere ( $M = 141$  ms) as compared to the right hemisphere ( $M = 135$  ms) for inverted stimuli,  $p = .02$ . Furthermore, the ANOVA showed that this interaction varied with Stimulus type,  $F(1,24) = 6.03$ ,  $p = .02$ ,  $\eta^2_p = .20$ . Post-hoc analyses revealed that the hemisphere effect for inverted stimuli was significant for sketched faces ( $p < .01$ ), but not for photographed ones ( $p = .90$ ). P100 peak latency in response to an inverted sketched face was shorter for the right (133 ms) than the left hemisphere (141 ms).

## Insensitivity to configural manipulations of emotional faces

### 2.2.2.3. N170 amplitude

The ANOVA showed greater N170 amplitude in response to inverted faces ( $M = -5\mu\text{V}$ ) than to upright stimuli ( $M = -4\mu\text{V}$ ),  $F(1,24) = 8.4$ ,  $p < .01$ ,  $\eta^2_p = .26$  (see Figure 3B&D). The only other significant effect was the Stimulus Type x Orientation interaction,  $F(1,24) = 69.19$ ,  $p < .001$ ,  $\eta^2_p = .74$ . Post-hoc analyses of sketch stimuli revealed greater amplitude for upright stimuli ( $M = -5.4\mu\text{V}$ ) as compared to inverted stimuli ( $M = -4.1\mu\text{V}$ ),  $p < .01$ . In contrast, for photographed stimuli, the amplitude was greater for inverted stimuli ( $M = -5.5\mu\text{V}$ ) than for upright stimuli ( $M = -2.3\mu\text{V}$ ),  $p < .001$ , see Figure 4B. The effect size of the absolute inversion effect for sketched faces (95% CI =  $[0.54\mu\text{V}; 2.22\mu\text{V}]$ ,  $d = 0.68$ ) was moderate in comparison to photographed faces (95% CI =  $[2.33\mu\text{V}; 4.14\mu\text{V}]$ ,  $d = 1.47$ ) for which the inversion effect was greater and more robust. Finally, post-hoc comparisons showed greater negativity for upright sketched faces as compared to upright photographed and inverted sketched faces,  $ps < .001$ .

### 2.2.2.4. N170 peak latency

The ANOVA showed greater N170 peak latency for sketches ( $M = 190$  ms) than for photographs ( $M = 182$ ms),  $F(1,24) = 19$ ,  $p = .001$ ,  $\eta^2_p = .44$ , and for inverted stimuli ( $M = 189$  ms) than upright stimuli ( $M = 183$  ms),  $F(1,24) = 35.64$ ,  $p < .001$ ,  $\eta^2_p = .60$ . The nonsignificant Stimulus type x Orientation interaction,  $F(1,24) < 1$ , n.s., suggests an additive effect of Stimulus type (10 ms increase for sketched faces) and Orientation (7 ms increase for inverted facial stimuli) on N170 peak latency (see Figure 4 A), along the lines of Sagiv & Bentin (2011). The effect size of the inversion effect (95% CI =  $[4$  ms;  $9$  ms],  $d = 1.18$ ) and of stimulus type (95% CI =  $[4$  ms;  $12$  ms],  $d = 0.86$ ) were of comparable magnitude. Finally, there was no main effect of Hemisphere, nor did Hemisphere interact with other factors.

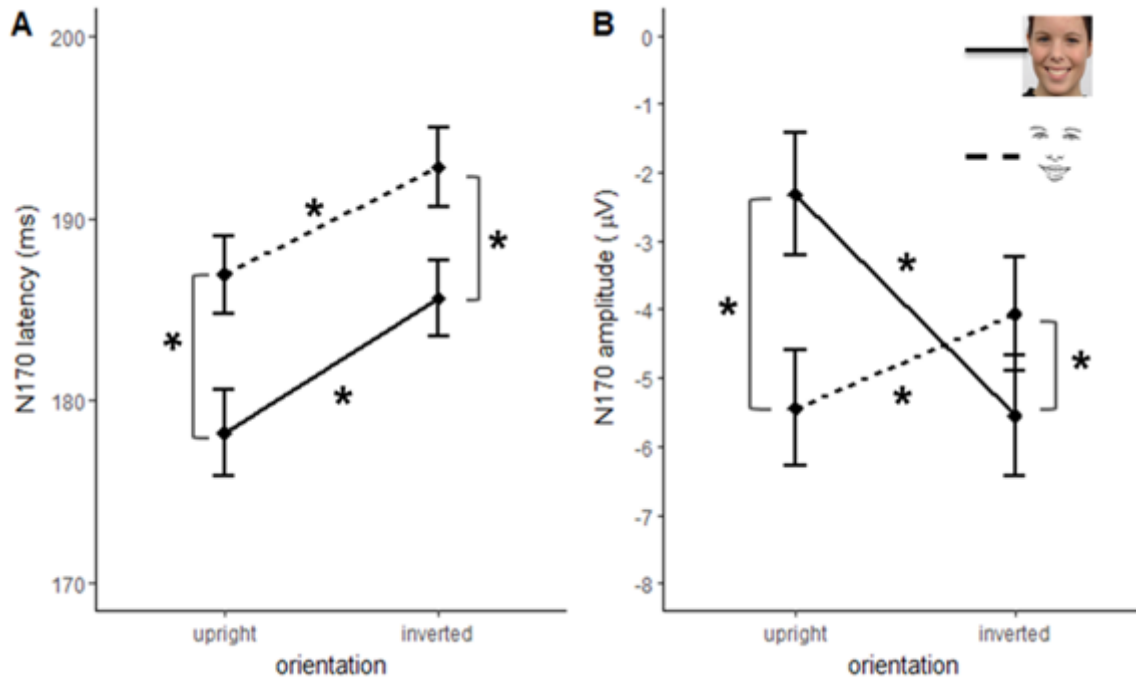


Figure 4. Mean N170 component (averaged for electrodes P7 & P8) (A) latencies, and (B) peak amplitude for the Stimulus type x Orientation interaction (\* $p < .05$ ). Error bars indicate standard error of the mean.

### 2.3. Discussion

In this experiment, we found no behavioral difference between sketched and photographed faces when presented in the upright orientation. This result suggests that diagnostic facial features provide sufficient information for recognizing facial expressions even without facial texture information. Indeed, with eye-tracking measure, 88% of gaze fixations are directed to the diagnostic features (e.g., the eyes, nose, mouth) across all the facial regions when judging facial expressions (Schurgin et al., 2014). Our results are in line with other studies showing that it is more difficult to process inverted faces than upright faces, resulting in longer reaction times and lower accuracy (Itier & Taylor, 2004; Minami et al., 2015). Regarding the face inversion effect, Hu performance was more degraded for sketched than for photographed faces, although reaction time increased similarly for both

## Insensitivity to configural manipulations of emotional faces

stimulus types. To our knowledge, this is the first time that the removal of facial information has been shown to reduce accuracy without increasing reaction time. The only face detection study comparing the face inversion effect between sketched and photographed faces reported no significant difference in accuracy between inverted sketched and photographed faces. These results suggest that the task influences the intake of visual information (Brosch, Pourtois, & Sander, 2010), and show that inverted sketched faces are somehow processed differently from photographed faces.

In addition to behavioral measures, we replicated the face inversion effect on N170 amplitude and peak latency for photographed faces, with delayed and increased N170 reflecting a disruption of configural processing (George, Evans, Fiori, Davidoff, & Renault, 1996; Bruno Rossion et al., 1999). Conversely, inverted sketched faces elicited reduced and delayed N170. These results are in line with Sagiv and Bentin's face detection study (2011) which also showed that the N170 was smaller and later when sketched faces were processed. Similar results were observed for inverted Mooney faces as compared to photographed faces (Itier et al., 2006). Inverted sketched faces evoke smaller and delayed N170 with respect to Mooney faces, suggesting that the recruitment of configural processing is modulated by stimulus quality as well as departure from the canonical upright orientation. Therefore, N170 peak latency revealed an additive effect for sketched and photographed faces leading to a greater N170, which is in line with Sagiv and Bentin (2001). Taken together, our results suggest that recognizing emotion from sketched faces relies on more part-based processing which leads to great and delayed N170 to achieve equivalent behavioral performance to upright photographed faces. Another interesting finding concerns P100 that exhibits a main effect of stimulus type with larger amplitudes in photographs than in sketches. This argues in favor of the association of this component with the detection of rich natural stimuli (possibly

Insensitivity to configural manipulations of emotional faces

related to differences in low-level visual cues, such as spatial frequencies), partially affected by orientation, and with inverted stimuli inducing additional neuroelectric activation.

### 3. Experiment 2

#### 3.1. Method and Procedure

The experimental procedure was similar to Experiment 1 with the same 25 participants performing Experiment 2 after Experiment 1. Only the facial stimuli differed with respect to Experiment 1, since facial stimuli were modified parametrically in terms of “degree of deconfiguration”. The task consisted of one block of 150 trials displaying upright sketch stimuli (5 expressors x 5 emotions x 6 degrees of deconfiguration). The deconfiguration consisted in a translation corresponding to the radius (in pixels) of a circle (centered on the feature) around which each face part was displaced. This parameter was manipulated from 0 to 25 pixels in five intervals. Each face part was moved randomly in a different direction to ensure that the configuration of the face was altered.

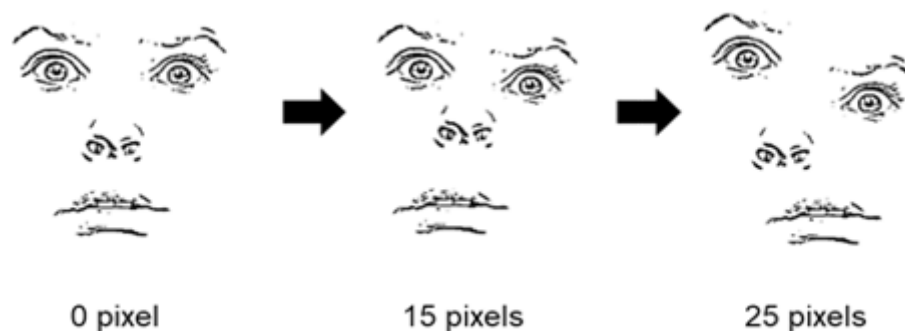


Figure 5: From left to right shows the 0, 15 and 25 pixels of deconfiguration for a fearful female trial.

#### 3.2. Data Analysis

Behavioral data and ERP data were analyzed as in Experiment 1. The only difference was in terms of experimental factors. We only considered a within-subject “degree of

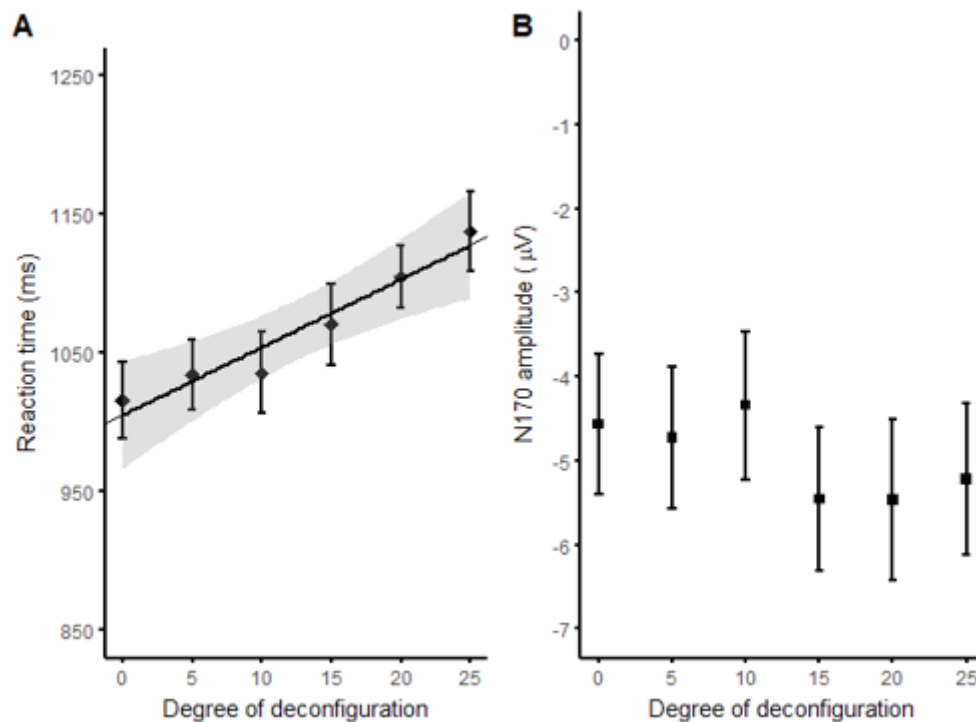
Insensitivity to configural manipulations of emotional faces

deconfiguration” factor for analyzing the behavioral data, and a supplementary Hemisphere within-subject factor was introduced for the ERP data analysis. Greenhouse-Geisser adjusted degrees of freedom were used when the sphericity assumption was not met ( $\epsilon$  values are given in this case).

### 3.3. Results

#### 3.3.1. Behavioral data

The ANOVA showed that the degree of deconfiguration did influence reaction time (RT) significantly,  $F(5,120) = 9.15$ ,  $p < .001$ ,  $\eta^2_p = .28$ , but did not affect response accuracy (Hu),  $F(5,120) < 1$ , n.s., which was .87 on average. Descriptively, RT increased with the degree of deconfiguration, hence, a trend analysis was conducted. The results showed that RT increased linearly,  $F(5, 120) = 48.70$ ,  $p < .001$ , by about  $5 \text{ ms} \pm 1.3 \text{ (SE)}$  per unit of degree of deconfiguration (see Figure 6 A).



### Insensitivity to configural manipulations of emotional faces

Figure 6: Mean effect of the degree of deconfiguration on (A) reaction time together with a linear trend fit (95% CI); and (B) N170 amplitude ( $\mu\text{V}$ ). Error bars indicate standard error of the mean.

#### 3.3.2. Neurophysiological data

ANOVA on P100 amplitude showed no effect of Hemisphere ( $F(1,24) = 1.33$ ,  $p = .26$ ,  $\eta^2_p = .053$ ) or of degree of deconfiguration ( $F(5,120) < 1$ , n.s.), and no significant interaction between the two factors ( $F(5,120) = 1.83$ ,  $p = .11$ ,  $\eta^2_p = .07$ ). ANOVA on P100 peak latency revealed a marginally significant effect of the degree of deconfiguration ( $F(5,120) = 2.20$ ,  $p = .058$ ,  $\eta^2_p = .08$ ), that combined a marginally significant linear effect (decreasing P100 peak latency with an increasing degree of deconfiguration),  $F(1,24) = 3.44$ ,  $p = .076$ , together with a significant U-shaped quadratic effect,  $F(1,24) = 6.71$ ,  $p = .016$  (see Figure 7). The linear and quadratic trends accounted for 33% and 49% of the variance in the effect of degree of deconfiguration, respectively. However, given that visual inspection clearly shows that the trend is linear from 0 to 20 pixels of deconfiguration, and the change comes afterwards, we performed a linear fit on the [0; 20 pix] data that turned out to be significant ( $t(24)=3.24$ ,  $p = .0035$ ) and accounted for 90% of the variance in the deconfiguration effect. There was no effect of Hemisphere ( $F(1,24) < 1$ , n.s.), nor did the two factors interact ( $F(5,120) < 1$ , n.s.)

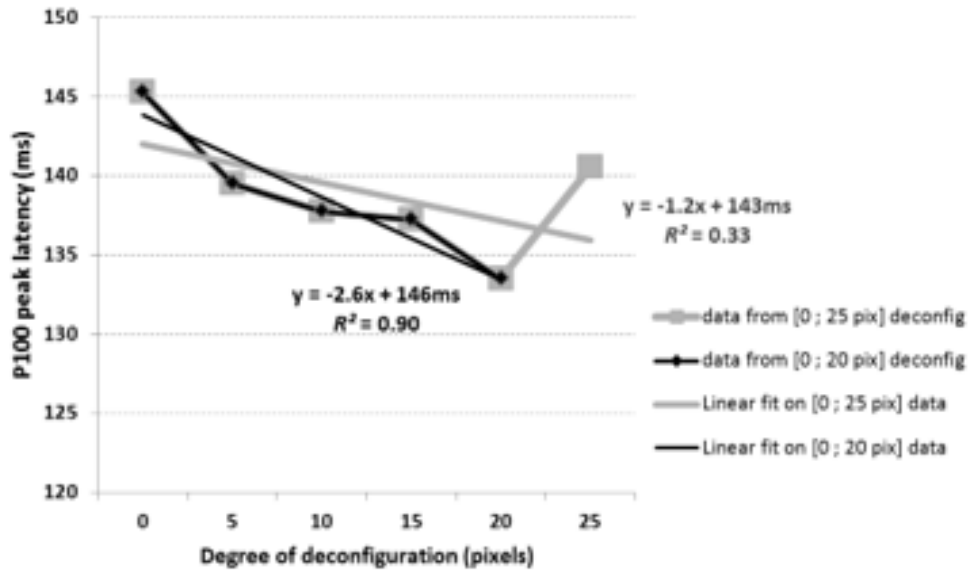


Figure 7. P100 peak latency as a function of degree of deconfiguration, with linear fits on the [0; 20 pix] and [0; 25 pix] data. Note that the slope is indicated per level of the deconfiguration factor ( $x = 5\text{ms}$  intervals).

ANOVA on N170 peak amplitude showed a marginally significant effect ( $F(5,120) = 2.22, p = .06, \eta^2_p = .08$ ) of Degrees of deconfiguration. Visual inspection of Figure 6 shows a peak amplitude dropout for degrees 15, 20, 25 ( $M = -5.38\mu\text{V}$ ) as compared to degrees 0, 5, 10 ( $M = -4.55\mu\text{V}$ ). In order to test whether this dropout was significant, we averaged the corresponding amplitude ( $\mu\text{V}$ ) values to compare both triplets (“0, 5, 10” vs “15, 20, 25”) and found a significant difference ( $t(24) = 2.82, p < .01$ ), see Figure 6. However, there was no effect of Hemisphere ( $F(1,24) = .31, p = .58, \eta^2_p = .02$ ) and no significant Hemisphere x Degree of deconfiguration interaction ( $F(5,120) = .29, p = .92, \eta^2_p = .01$ ). ANOVA analysis on N1700 peak latency revealed neither a main effect of Hemisphere ( $F(1,24) = 4.15, p = .53, \eta^2_p = .15$ ) nor of degree of deconfiguration ( $F(5,120) = 1.66, p = .15, \eta^2_p = .06$ ), and no interaction between both factors ( $F(5,120) < 1, \text{n.s.}$ )

### 3.4. Discussion

## Insensitivity to configural manipulations of emotional faces

This experiment showed that randomly changing spatial distances between diagnostic features did not affect accuracy, although it did result in a linear increase of processing time at the behavioral level (RT). In contrast, the effect of our parametric manipulation of the sketch stimuli was nonlinear on P100 peak latency (quadratic trend) and N170 peak amplitude (dropout at the median degree of deconfiguration). Halit et al. (2000) manipulated the distance between facial feature with “stretched faces” on photographs and found greater and delayed P100 but no effect on N170. They suggested that N170 was affected only when typical faces are presented. From Experiment 1, we observed that inverted sketched faces failed to elicit N170s similar to inverted photographed faces, suggesting that the configural information was somehow impaired through the sketch rendering/face contour removal procedure. These results lead us to conclude that sketched faces recruit more part-based processing to maintain emotion recognition performance, leading to a deeper recruitment of neuron processing (greater N170 amplitude) at greater spatial deconfiguration in Experiment 2, similar to the greater N170 amplitude for upright sketched faces compared to upright photographed faces in Experiment 1.

## 4. General Discussion

Current EEG studies investigate the effect of several configural manipulations: (i) the removal of facial texture (and face contour) from photographed faces to sketched faces; (ii) face inversion on both sketched and photographed faces; (iii) changed spatial distances of diagnostic features on sketched faces in the specific context of emotional expression recognition. Finally, to organize our findings, we will discuss the results within a comprehensive model that accounts for behavioral and neuroelectrical data.

Our results extend the findings of Sagiv & Bentin (2001) who showed that the N170 face inversion effect was observed only for photographed faces but not for sketches. The task in the study of Sagiv & Bentin (2001) was to count mentally presented objects, i.e., flowers,

### Insensitivity to configural manipulations of emotional faces

while passively viewing photographed and sketched faces in upright and inverted orientation (in their second experiment). Contrary to this paradigm, the present study used an explicit emotion recognition task increasing the salience of face features. Still, we obtained similar N170 modulation, supporting that N170 is robust respond to facial stimuli and not to task differences (Flevaris et al., 2008; Goffaux et al., 2003). This is in line with existing studies showing that photographed faces elicit configural processing, whereas the removal of facial texture information alters configural information in a way that reinforces part-based processing, as observed for sketched faces (Meinhardt-Injac et al., 2013; Sagiv & Bentin, 2001; Zhao et al., 2016). In contrast to N170 amplitudes, P100 amplitudes showed no interaction between stimulus type (photograph vs. sketch) and stimulus orientation (upright vs. inverted) in facial expression processing. P100 peak amplitude was greater when faces are inverted instead of upright, and for photographed instead of sketched faces. The effect of stimulus type and orientation was additive, with a greater effect for stimulus type on P100 peak amplitude. The efficient extraction of configural information might occur at early face processing during the P100 time window. These findings support that the removal of facial texture reduces configural processing on sketched faces. On the other hand, the greater P100 amplitude for photographs is consistent with the visual system being attuned to rich stimulation (in terms of low-level properties). Nevertheless, the presumed impaired configural processing on sketch stimuli was not associated with low behavioral performance. The participants scored high accuracy and had similar RTs on both photographed and sketched faces when the faces were presented in the upright orientation. As hypothesized, the participants might exploit featural information on sketched faces, resulting in a similar behavior outcome to photographed faces.

We found an interaction between face orientation (inversion effect) and stimulus type (photographs vs. sketches) corresponding to greater N170 amplitudes for inverted faces than

### Insensitivity to configural manipulations of emotional faces

upright faces in photographed faces, and the opposite in sketches. Some studies have investigated the interaction between face inversion and facial structural characteristics. For example, a face stimulus designed with non-face objects (i.e., Arcimboldo paintings) led to similar N170 amplitudes for Arcimboldo paintings and objects (Caharel et al., 2013) when displayed upside-down. Face inversion effects in other studies resulted in smaller N170 amplitudes for sketched faces compared to photographed faces (Sagiv & Bentin, 2001), and for high-spatial frequency faces compared with low-spatial frequency faces (Flevaris et al., 2008; Goffaux et al., 2003). These findings suggest that configural information is somewhat reduced in non-photographed faces displayed in the canonical upright orientation. Therefore, the enhancement of N170 for non-photographed faces might reflect sensitivity to the configural organization of inner facial components and the decoding of facial structure. Moreover, since the facial contour was removed on sketched faces, the processing of grouping the inner facial features may have enhanced N170 amplitudes, as suggested by Bentin et al., (2006). Alternatively, upright sketched faces were deprived of configural information. Taken together, these results suggest that a different face processing strategy is required to decode facial expressions efficiently from the different types of stimuli. Despite the removal of facial texture on sketches influencing facial expression processing, the presence of diagnostic facial features allows individuals to perform the task.

Our second experiment provided additional evidence in favor of more part-based processing of sketched faces in high recognition accuracy across different degrees of deconfiguration, accompanied with a linear increase of processing time (RTs) as stimulus configural distortion increased. Additionally, P100 and N170 amplitudes did not vary nonlinearly in response to linear parametric deconfiguration. Thanks to additional processing time, recognition accuracy of facial expressions from sketched faces may not be affected by spatial deconfiguration of the inner features of the face, at least within our range of

Insensitivity to configural manipulations of emotional faces

transformation (the maximum 25 pixels corresponded to 6.25% of stimulus size). This suggests that configural information is reduced by removing facial texture and retaining the most diagnostic features for emotion recognition.

#### **4.1. Summary: an attempt to model the task**

Gathering the evidence presented in this article, a feedforward accumulation model may be helpful to explain how visual input characteristics influence behavioral and neurophysiological data (amplitudes and latencies) during emotion recognition (see Figure 8). In such a model, visual features like shapes, textures and configuration are processed through specialized systems (detectors, recognizers), all of which are used to trigger the response when the threshold is reached. Using this framework, we focus on the hypothetical relationship between several observations: the existence of quick and ample electrical components (P100, N170) under certain conditions, and an accelerated behavioral response downstream. This also allows consideration of the balance between both components as a means to understand which systems are preferentially involved. The most striking effect described here is that the presentation of upright photographed faces combines a rapid behavioral response with a balance of P100 and N170 magnitudes in favor of the first component. This is compatible with the existence of highly oriented systems, like whole emotional face template recognizers, for the processing of these emotional stimuli. The second point is that the presentation of inverted faces retains an ample P100 but associates a deep N170 with it. This is consistent with the assumption that the generators underlying P100 are triggered by simple face detection (i.e., face template matching), but also that additional systems devoted to analyzing face structure are required later on. As mentioned above, there is considerable experimental evidence that N170 is strongly influenced by the load imposed on featural processing (Bentin et al., 2006; Itier et al., 2006; Neath & Itier, 2014). The new data provided by the sketches demonstrate the significant interference of such stimuli, and

# Insensitivity to configural manipulations of emotional faces

although they are associated with good behavioral performance, different parts of the model were engaged. First of all, P100 amplitude shows that early processing is impaired, presumably reducing the ability to detect a face, to recognize it by comparison with whole face templates, and even to infer its configuration. In the present model, this will, in cascade, favor the balance toward N170, but with marginal modulations by configuration and longer latencies. Indeed, slow and possibly sequential part-based processing feeds the accumulator with only partial and weak evidence.

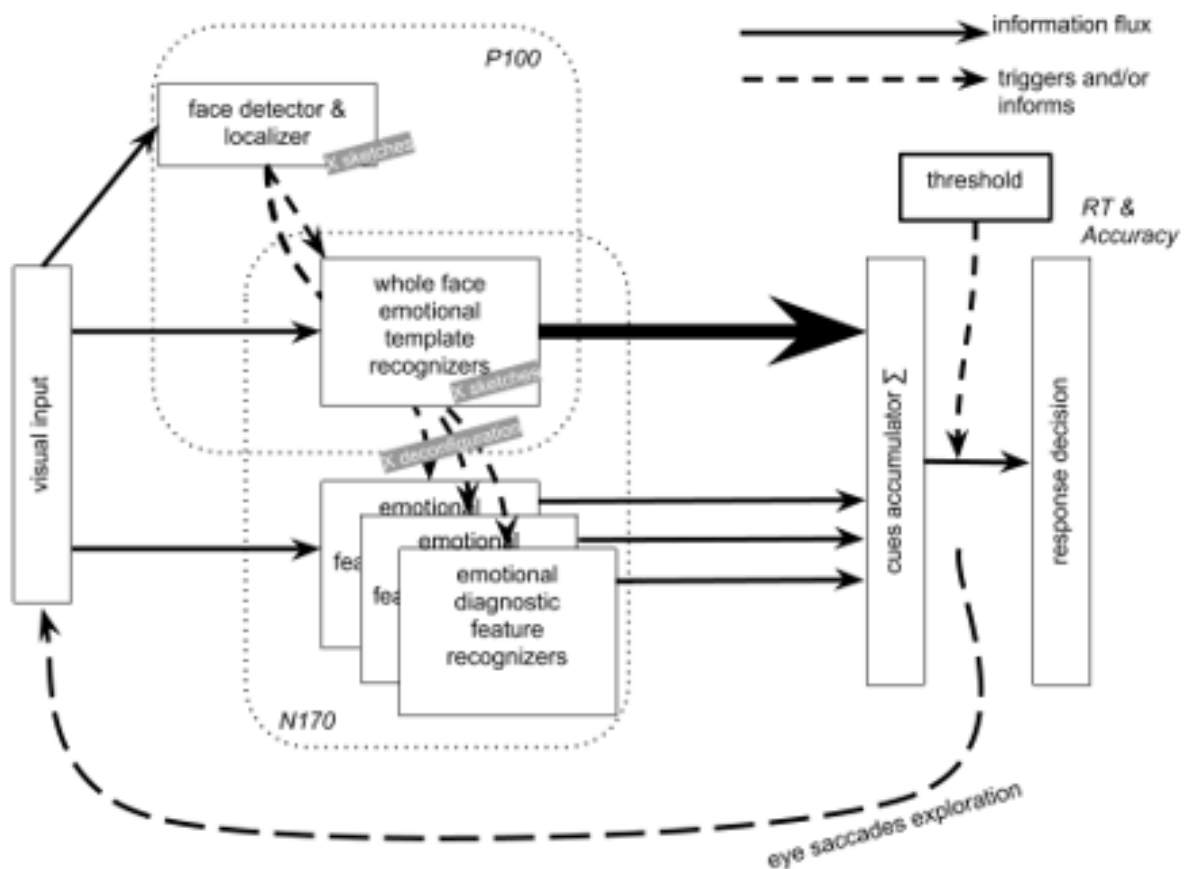


Figure 8. This ad hoc model is an attempt to conceptualize the present emotion recognition tasks with their different visual and configural manipulations and to integrate the findings from both experiments. See text for details.

## Insensitivity to configural manipulations of emotional faces

Our model, illustrated in Figure 8, shows that visual input on the left is processed and feeds the response decision system on the right through the progressive accumulation of cues that allows a response when a certain threshold is reached (this accounts for accuracy and reaction times). This processing involves a cascade of several brain systems (such as OFA, FFA, amygdala, see Pitcher et al., 2011) and generates successive ERP components (here, P100 and N170). First, the quickest system to respond to visual input is a face detector that directs attention towards a face stimulus and may inform the spatial coordinate system (coding the XYZ coordinates where parts/subparts are placed). When activated with facial stimuli, this system in turn activates template recognizers and feeds them with some indications on the visual areas to consider, helping them working faster. The second system is a set of whole face templates that allows the immediate recognition of emotional expressions when presented in a classical manner and feeds the accumulator with powerful cues. Finally, emotional diagnostic feature recognizers help identify emotions from specific face parts to further feed the accumulator when evidence is partial or weak. In this model, each recognizer is orientation-sensitive, being tuned to the canonical upright orientation of the face (along the lines of Jemel, Coutya, Langer, & Roy, 2009; Perrett, Oram, & Ashbridge, 1998). If the accumulation of evidence is too slow due to a strong departure of the stimulus from the canonical orientation, a reorienting processes will be recruited afterwards to match the template (such as top-down mental rotation processes occur processes occur between 340 and 700 ms, see Milivojevic, Hamm, & Corballis, 2011).

Alternatively, we may adopt the stance of Perrett et al. (1998) without invoking reorientation processes: “there is no need to pre-process an object to realise its likely orientation so that a mental rotation can take place subsequently. The system just accepts evidence in proportion to the activity of the relevant visually responsive cells in temporal cortex and elsewhere. Cells responsive to the face in temporal cortex will have outputs to

## Insensitivity to configural manipulations of emotional faces

brain systems capable of accessing semantic associations for faces (including verbal labels in the case of humans) and activating appropriate behavioural outputs. These associations will be more likely to be retrieved when face stimuli are presented in a frontal view, upright orientation and at a normal size, because under these conditions there will be more activity from cells in the temporal cortex already tuned to faces. Mental rotation has been invoked as an additional process to cope with unusual views. By contrast it is argued here that both typical and unusual views can be recognised using viewer-centred cells which operate in the same manner; unusual views take longer simply because there is less machinery dedicated to their processing." (p.139). Depending on the visual input (c.f. our experimental manipulations) this processing model behaves differently. When upright photographs are presented, the face detector and the whole face template recognizers are activated, creating a rapid and robust P100, and the output threshold is reached quickly. Note that this does not imply a fixed output/decision threshold. Along the lines of Palmeri, Schall, & Logan (2015), there are least four different reasons that may lead to a faster decision due to better visual input quality (i.e., closer to what the system is tuned to): a) earlier signal accumulation (accumulation onset); b) faster signal accumulation (accumulation growth rate); c) higher baseline level; and d) decreased threshold at which the accumulator sends its output. Figure 9 illustrates how the accumulation of evidence also applies locally to the N170 generator, and how it combines both N170 peak amplitude and latency.

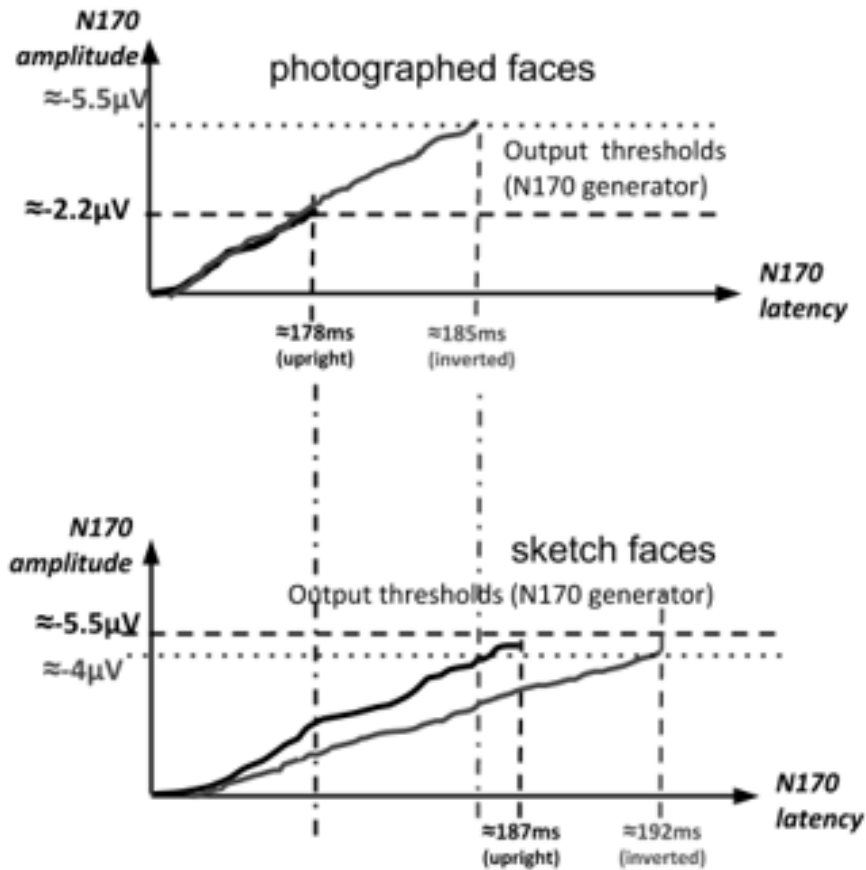


Figure 9. Theoretical underlying accumulation of evidence in the N170 generator accounting for the data observed in Experiment 1. Here, the N170 amplitude is supposedly quantitatively related to the output threshold of its generator, which in turn depends on its default stimulus tuning (i.e., rich upright stimuli). Upper panel: similar growth rate for photographed faces whether upright or inverted. The default threshold level is low when the stimulus quality is closer to what the system is tuned to. Here, inverted photographed faces increase N170 output latency due to an increase of the N170 generator response threshold. Lower panel: the threshold for sketched faces stays high for the reasons mentioned previously. In addition, the growth rate decreases with impoverished quality of the stimulus, which in turn increases the N170 latency.

## Insensitivity to configural manipulations of emotional faces

Accordingly, a rapid and robust P100 may be thought to modulate each of these parameters. When photographs are inverted, the face detector triggers a) whole face template recognizers that fail to process information quickly, but also b) feature recognizers, which create stronger N170 activity, due to an increase in the N170 output threshold. These activities are slightly delayed as more evidence needs to be accumulated (see Figure 9). In the case of sketches, the face detector is activated weakly and does not prime the other systems with specific spatial localizations of low-level face cues, so P100 remains low. Thus, in the case of upright sketches, single emotional recognizers may detect face parts and progressively accumulate evidence in favor of a specific emotion. This process leads to a moderate and slow N170, and to lengthy response times. In the case of inverted sketches, the accumulation of evidence growth rate decreases even more (see Figure 9) which increases N170 latency with a threshold kept high due to the poor stimulus input. Data from Experiment 2 can also be interpreted with the same accumulation of evidence framework. The N170 generator seems to tolerate a given amount of deconfiguration, up to a given degree (15 pix) of deconfiguration, from which the output threshold is increased (see peak amplitude dropout in Figure 6B). Accordingly, deconfiguration of the sketch stimuli contributes a marginal additional cost to the diagnostic feature recognizers and results in a decreased growth rate of evidence accumulation such that the place where diagnostic cues are situated does not match the default spatial configuration to which the system is tuned. Last, we restate here that this model accounts only for the early brain processes participating in the decision, and to guide further visual exploration (saccades orientation) to fetch new information to make a decision.

### 4.2. Perspectives

Twenty years after the seminal article by Perrett et al (1998), “accumulation of evidence” appears to be the relevant concept for understanding how altered configural

### Insensitivity to configural manipulations of emotional faces

information (stimulus impoverishment or spatial deconfiguration) or processing (stimulus inversion) affects emotion recognition. Our findings not only provide new empirical data along these lines among healthy participants, but also new methodological perspectives that can be applied in the field of mental disorders where face processing is supposedly altered and concomitant to the pathology. The facial components of our sketch stimuli are separable and easily superimposed onto backgrounds with similar textures, facilitating the design of many classical gestalt paradigms (embedded figures, contour matching, etc.). To exemplify the relevance of such stimulus manipulation, another study using EEG embedded neutral sketched faces with sketched tree branches to determine whether patients with schizophrenia can detect the face (Maher, Mashhoon, Ekstrom, Lukas, & Chen, 2016). They found that N170 amplitudes were greater for viewing faces than trees in control subjects, but no difference between viewing trees and faces was found in patients with schizophrenia. The results of our study provide additional clarification on the fact that some degraded stimuli engage featural processing that require integration in a gestalt representation, a process akin to contextual processing that may be impaired in these mental disorders. Last, this knowledge is of real theoretical importance in understanding the impact of cognitive training/remediation techniques for emotional recognition based on the use of non-natural or simplified stimuli

## 5. References

- Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12(2), 169–177. [https://doi.org/10.1016/S0959-4388\(02\)00301-X](https://doi.org/10.1016/S0959-4388(02)00301-X)
- Anaki, D., Zion-Golumbic, E., & Bentin, S. (2007). Electrophysiological neural mechanisms for detection, configural analysis and recognition of faces. *Neuroimage*, 37(4), 1407–1416.
- Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, 25(3), 281–316.
- Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research*, 17(3), 613–620.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8(6), 551–565.
- Bentin, S., Golland, Y., Flevakis, A., Robertson, L. C., & Moscovitch, M. (2006). Processing the trees and the forest during initial stages of face perception: Electrophysiological evidence. *Journal of Cognitive Neuroscience*, 18(8), 1406–1421.
- Bombardi, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion recognition: The role of featural and configural face information. *The Quarterly Journal of Experimental Psychology*, 66(12), 2426–2442. <https://doi.org/10.1080/17470218.2013.789065>
- Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cogn. Emot.*, 24(3), 377–400.

# Insensitivity to configural manipulations of emotional faces

- Caharel, S., Collet, K., & Rossion, B. (2015). The early visual encoding of a face (N170) is viewpoint-dependent: a parametric ERP-adaptation study. *Biol. Psychol.*, 106, 18–27.
- Caharel, S., Leleu, A., Bernard, C., Viggiano, M.-P., Lalonde, R., & Rebaï, M. (2013). Early holistic face-like processing of Arcimboldo paintings in the right occipito-temporal cortex: evidence from the N170 ERP component. *International Journal of Psychophysiology*, 90(2), 157–164.
- Eng, Z. H. D., Yick, Y. Y., Guo, Y., Xu, H., Reiner, M., Cham, T. J., & Chen, S. H. A. (2017). 3D faces are recognized more accurately and faster than 2D faces, but with similar inversion effects. *Vision Research*, 138, 78–85.  
<https://doi.org/10.1016/j.visres.2017.06.004>
- Feng, W., Martinez, A., Pitts, M., Luo, Y.-J., & Hillyard, S. A. (2012). Spatial attention modulates early face processing. *Neuropsychologia*, 50(14), 3461–3468.
- Flevaris, A. V., Robertson, L. C., & Bentin, S. (2008). Using spatial frequency scales for processing face features and face configuration: an ERP analysis. *Brain Research*, 1194, 100–109.
- George, N., Evans, J., Fiori, N., Davidoff, J., & Renault, B. (1996). Brain events related to normal and moderately scrambled faces. *Cognitive Brain Research*, 4(2), 65–76.
- Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, 22(6), 1094–1118.  
<https://doi.org/10.1080/02699930701626582>
- Goffaux, V., Jemel, B., Jacques, C., Rossion, B., & Schyns, P. G. (2003). ERP evidence for task modulations on face perceptual processing at different spatial scales. *Cognitive Science*, 27(2), 313–325.

# Insensitivity to configural manipulations of emotional faces

- Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 1023.
- Halit, H., de Haan, M., & Johnson, M. H. (2000). Modulation of event-related potentials by prototypical and atypical faces. *Neuroreport*, 11(9), 1871–1875.
- Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, 33(1), 98–106.
- Herrmann, M., Ehlis, A.-C., Ellgring, H., & Fallgatter, A. (2005). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *Journal of Neural Transmission*, 112(8), 1073–1081.
- Itier, R. J., Alain, C., Sedore, K., & McIntosh, A. R. (2007). Early face processing specificity: It's in the eyes! *Journal of Cognitive Neuroscience*, 19(11), 1815–1826.
- Itier, R. J., Latinus, M., & Taylor, M. J. (2006). Face, eye and object early processing: what is the face specificity? *Neuroimage*, 29(2), 667–676.
- Itier, R. J., & Taylor, M. J. (2004). Face recognition memory and configural processing: a developmental ERP study using upright, inverted, and contrast-reversed faces. *Journal of Cognitive Neuroscience*, 16(3), 487–502.
- Jemel, B., Coutya, J., Langer, C., & Roy, S. (2009). From upright to upside-down presentation: A spatio-temporal ERP study of the parametric effect of rotation on face and house processing. *BMC Neuroscience*, 10(1), 100. <https://doi.org/10.1186/1471-2202-10-100>
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>

# Insensitivity to configural manipulations of emotional faces

- Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception: Early visual experience and face processing. *Nature*, 410(6831), 890.
- Leder, H. (1999). Matching person identity from facial line drawings. *Perception*, 28(9), 1171–1175.
- Leder, Helmut, & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology: Section A*, 53(2), 513–536.
- Maher, S., Mashhoon, Y., Ekstrom, T., Lukas, S., & Chen, Y. (2016). Deficient cortical face-sensitive N170 responses and basic visual processing in schizophrenia. *Schizophrenia Research*, 170(1), 87–94.
- Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255–260.
- McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review*, 16(5), 778–797. <https://doi.org/10.3758/PBR.16.5.778>
- Meaux, E., & Vuilleumier, P. (2016). Facing mixed emotions: Analytic and holistic perception of facial emotion expressions engages separate brain networks. *NeuroImage*, 141, 154–173. <https://doi.org/10.1016/j.neuroimage.2016.07.004>
- Meinhardt-Injac, B., Persike, M., & Meinhardt, G. (2013). Holistic Face Processing is Induced by Shape and Texture. *Perception*, 42(7), 716–732. <https://doi.org/10.1068/p7462>
- Milivojevic, B., Hamm, J. P., & Corballis, M. C. (2011). About turn: How object orientation affects categorisation and mental rotation. *Neuropsychologia*, 49(13), 3758–3767. <https://doi.org/10.1016/j.neuropsychologia.2011.09.034>

# Insensitivity to configural manipulations of emotional faces

- Minami, T., Nakajima, K., Changvisommid, L., & Nakauchi, S. (2015). The effects of facial color and inversion on the N170 event-related potential (ERP) component. *Neuroscience*, 311, 341–348.
- Namdar, G., Avidan, G., & Ganel, T. (2015). Effects of configural processing on the perceptual spatial resolution for face features. *Cortex*, 72, 115–123.
- Neath, K. N., & Itier, R. J. (2014). Facial expression discrimination varies with presentation time but not with fixation on features: a backward masking study using eye-tracking. *Cognition & Emotion*, 28(1), 115–131.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Palmeri, T. J., & Cottrell, G. W. (2010). Modeling Perceptual Expertise. In I. Gauthier, T. Michael, & B. Daniel, *Perceptual Expertise: Bridging Brain and Behavior* (pp. 197–244). Oxford University Press.
- Palmeri, T. J., Schall, J. D., & Logan, G. D. (2015). Neurocognitive modeling of perceptual decision making. In J. R. Busemeyer, W. Zheng, J. T. Townsend, & A. Eidels, *The Oxford Handbook of Computational and Mathematical Psychology* (pp. 320–340). Oxford University Press.
- Perrett, D. ., Oram, M. ., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition*, 67(1–2), 111–145. [https://doi.org/10.1016/S0010-0277\(98\)00015-8](https://doi.org/10.1016/S0010-0277(98)00015-8)
- Piepers, D. W., & Robbins, R. A. (2012). A Review and Clarification of the Terms “holistic,” “configural,” and “relational” in the Face Perception Literature. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00559>

# Insensitivity to configural manipulations of emotional faces

- Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, 209(4), 481–493. <https://doi.org/10.1007/s00221-011-2579-1>
- Regan, D. (1989). *Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine*.
- Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research*, 51(3), 333–342.
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects. *Neuroreport*, 11(1), 69–72.
- Rossion, Bruno, Delvenne, J.-F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., & Guérit, J.-M. (1999). Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biological Psychology*, 50(3), 173–189.
- Sadeh, B., & Yovel, G. (2010). Why is the N170 enhanced for inverted faces? An ERP competition experiment. *Neuroimage*, 53(2), 782–789.
- Sagiv, N., & Bentin, S. (2001). Structural encoding of human and schematic faces: holistic and part-based processes. *Journal of Cognitive Neuroscience*, 13(7), 937–951.
- Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision*, 14(13), 14–14.
- Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and Decoding Facial Expressions. *Psychological Science*, 16(3), 184–189. <https://doi.org/10.1111/j.0956-7976.2005.00801.x>
- Tanaka, J. W., Kaiser, M. D., Butler, S., & Le Grand, R. (2012). Mixed emotions: Holistic and analytic perception of facial expressions. *Cognition & Emotion*, 26(6), 961–977. <https://doi.org/10.1080/02699931.2011.630933>

- Tzschaschel, E., Persike, M., & Meinhardt-Injac, B. (2014). The effect of texture on face identification and configural information processing. *Psihologija*, 47(4), 433–447. <https://doi.org/10.2298/PSI1404433T>
- Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, 17(1), 3–28.
- Wang, H., & Fu, S. (2018). Spatial attention modulates the temporal sequence of hemispheric asymmetry in configural and featural face processing. *Neuropsychologia*, 111, 269–275. <https://doi.org/10.1016/j.neuropsychologia.2018.01.029>
- Wang, H., Sun, P., Ip, C., Zhao, X., & Fu, S. (2015). Configural and featural face processing are differently modulated by attentional resources at early stages: An event-related potential study with rapid serial visual presentation. *Brain Research*, 1602, 75–84.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1), 141.
- Zhao, M., Bülthoff, H. H., & Bülthoff, I. (2016). A shape-based account for holistic face processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(4), 584–597. <https://doi.org/10.1037/xlm0000185>

## 6. Supplementary material

The Practice session included two sub-sessions: (1) a key learning session to ensure the participant was able to correctly associate a given response with a specific emotional state; (2) the facial emotion recognition session. The “D”, “F”, “H”, “J” and “K” keys of an AZERTY keyboard were re-labeled with stickers as follows: “C” (for “Colère”, meaning anger in French); “P” (for “Peur” = fear); “N” (for “Neutre” = neutrality), “T” (for “Tristesse” = sadness), and “J” (for “Joie” = happiness). The key learning session consisted of a minimum of 30 trials with one emotional state word displayed randomly in the center of a screen to which participants responded by pressing the corresponding key. After each five-trial block, the average response accuracy was displayed (in %). This key learning session ended when participants reached 100% accuracy over five successive blocks. Then, the facial emotion recognition practice session began. There were three blocks, i) inverted and upright sketched faces; ii) inverted and upright photographed faces; and iii) sketched faces in deconfiguration. Block order was randomized. After each block of five trials, average response accuracy (in %) was displayed. The practice trials consisted of 40 images for the sketched and photographed face blocks: 2 Expressors x 2 Orientation (inverted vs. upright) x 2 Stimulus type (sketch vs. photograph) x 5 Facial expressions. It consisted of 60 images for the sketch deconfiguration block: 2 Expressors x 6 Degree of deconfiguration x 5 Facial expressions. The stimuli used in the practice sessions were excluded in the experimental session. Participants obtaining less than 75% accuracy during the practice session did not take part in the experimental session.



## General discussion

This thesis aimed to solve the fundamental question of how a facial expression is decoded in relation to the underlying neural mechanisms of face processing. Facial features are known to be important for recognizing an emotion, but the influence of configural processing on the processing of facial expressions was only addressed by a few studies. In order to investigate the contribution of configural and featural information for emotion recognition, we created a new set of sketch face stimuli that retains only the diagnostic facial features such as eyes, nose and mouth by transforming photographed face stimuli from the Radboud Face Database ([Langner et al., 2010](#)). This sketch stimuli set expressed four basic emotions (e.g. angry, sad, happy and fearful) and neutral expression. We compared the performance of participants on sketch faces to their photographed counterparts in a emotion recognition task with behaviour, Eye-tracking, and EEG measurements. Four experiments were conducted on healthy participants to examine their visual processes using the two emotional face stimuli (sketch and photographed faces). In addition, an EEG pilot study on two participants with the neurodevelopmental disorder schizophrenia was performed (see Appendix [5.1](#)).

At first, a validation of the new set of sketch face stimuli was conducted with behavioral measurements. The study was used to confirm that the created sketch stimuli, containing only diagnostic facial features and no further visual information (e.g., hair, shadow or wrinkles), enable the categorization as emotional faces (Experiment 1). Based on these results, the best-recognized sketch faces were chosen (ten face stimuli

for each emotion) for further eye-tracking and EEG studies. In the second experiment, eye-tracking measurements were used to compare the allocation of the visual attention and to explore similarities and differences between sketch and their photographed counterpart faces. The duration and location of the mean fixation, first and second fixations were examined during an emotion recognition task for both stimulus types (Experiment 2). Thirdly, EEG measures were performed to understand the impact of the configural information further using the inverted sketch and photographed counterpart faces with a focus on the examination of the P100 and N170 components (Experiment 3). In Experiment 4, we evaluated the importance of the face configuration for recognizing facial expressions by manipulating the distances between facial features of the sketch stimuli. The influence of this face deconfiguration was investigated with EEG measurements and the ERP components as well as the behavioral performance evaluated.

In sum, this thesis correlates the results of fixation patterns and underlying cognitive processing indicated by the ERPs (P100 and N170) with behavioral performances to better understand the visual processing strategy for correct emotion recognition. Within this final chapter, the results from all studies will be summarized and discussed with regard to their implications to our understanding of early facial emotion processing.

## **4.1 Behavioral performance**

### **4.1.1 Sketch faces**

Based on the assumption that each facial expression has its typical physiological characteristic, sketch faces should convey sufficient expressive information when the most important diagnostic features are presented. Although sketch faces reduce the effectiveness of representing information in face reconstruction in comparison to

photographed faces, the edge-based information seems to emphasize the contour of facial features strongly. The results for example in an easier recognition of happy expressions as it has a unique feature across all emotions, which is a curved smiling mouth. In Experiment 1, happy expressions were processed quicker and recognized more accurate than other emotions. This unique feature is known to trigger part-based processing for the recognition of the happy expression. This was shown for example in a study of [Calvo et al. \(2014\)](#) in which they further revealed that the process occurred early at around 150-180 ms. The importance of those salient features was supported by a study of [Rossion \(2013\)](#) in which they argued that the face perception could be activated automatically even only one part is available to the visual system when this part is a highly distinctive diagnostic feature. Our sketch faces contain the most expressive diagnostic features of a face. Therefore, the mouth feature might evoke part-based processing and would account for quicker behavioral recognition of the happy expression.

Interestingly participants further recognized well the fearful expression on sketch faces. One explanation could be that sketch faces enhance the contrast of the sclera and pupil in the eye region. [Whalen \(2004\)](#) has demonstrated that the white part of fearful eyes (i.e., sclera) cause an increased response in the left ventral of the amygdala in comparison to the eyes of the happy expression. This means that the wide-open eyes with a big contrast in black and white would also explain the better recognition of the fearful expression in comparison to angry, sad and neutral expressions. An example of how important the eye region for emotional recognition is can be found for persons who have bilateral amygdala damage. They show limited processing of information from the eye region causing poor performance in judging emotions. ([Adolphs et al., 2005](#)).

Sad and angry expressions appear to be the most difficult to recognize in comparison to other emotions (in Experiment 1 & 2). [Calvo and Nummenmaa \(2008\)](#) have found that sad faces were often misclassified as angry faces. The low accuracy of sad and angry expressions is also reported in other studies ([De Sonnaville et al., 2002](#); [Wegrzyn et al., 2017](#)), suggesting the amount of overlapping facial characteristics (e.g., AUs) between these emotions increased the ambiguity. It further seems that the

accuracy of recognizing angry and sad expressions increases when the intensity of the facial expression also increases. This means, on the other hand, that the removal of facial texture and shape information might weaken the expressive intensity on our sketch faces. In addition to the facial stimuli information in the original RaFD, [Langner et al. \(2010\)](#) also examined the unbiased hit rates between female and male adult models (stimuli) for participants. They found that female models had a higher accuracy than male models, especially for happy expression ([Langner et al., 2010](#)). As the used facial stimuli in our study were equalized in gender, the poor recognition performance of the angry expression in our studies might be related to the gender effect. Because angry expressions are better recognized on male than female based on the socialization of aggressiveness might involve with our experience ([Palermo and Coltheart, 2004](#); [Wegrzyn et al., 2017](#)). Future studies should investigate the effect of intensity of facial expressions and the difference between gender on models.

#### **4.1.2 Comparison between sketch and photographed faces**

The direct comparison between emotional sketch and photographed faces showed comparable performances in our experiments (Experiment 2 & 3). Although sketch stimuli contain less facial textural information (such as shadows, hair, color, etc.), the presence of the diagnostic facial features provides sufficient expressive information for the perception of facial expressions. These results support the general hypothesis that the diagnostic facial features (e.g., eyes, nose, mouth) play a crucial role in emotion recognition ([Schurgin et al., 2014](#); [Bombari et al., 2013](#)). The behavioral performance in recognizing happy faces produced a high accuracy and short reaction times across all experiments conducted in this thesis. These results are in line with the aforementioned happy faces having the unique upward mouth characteristic and being the only positive expression ([Calvo and Nummenmaa, 2008](#); [Langner et al., 2010](#)). On both sketch and photographed faces, we also found angry and sad expressions were poorly recognized. In a comparison study using scrambled, blurred and intact photographed faces, it was illustrated that the recognition of anger and sadness was significantly affected by scrambled face effect ([Bombari et al., 2013](#)). This result again suggests that sad and

angry faces share many facial characteristics, particularly the upper facial features, resulting in a longer processing time required to gather more global information (link to configural processing). Additionally, the recognition of anger is very subtle, and the space between the eyes and the mouth is very similar to sad face (Wegrzyn et al., 2017). Hence, angry expression can be confused as another emotion, such as sad.

Although for both stimuli the emotion recognition occurred with high accuracy, there are differences regarding the reaction times. Reaction times for recognizing neutral expressions was close to happy expressions with photographed faces. However, we found a significantly longer reaction time for the neutral expression on sketch faces in comparison to the happy expression. The fact that photographed faces provide rich facial texture information (shadow, depth, and color, etc.), as well as non-expressive information (hair, ear, face contour, etc.) apart from the diagnostic features, probably results in a faster recognition of neutral expressions for photographed faces in comparison to sketch faces. Therefore the non-emotional related visual information is important for the identification of the neutral expression. This importance is supported by the fact that there are also no action units defined for the description of a neutral expression. Thus, participants have a longer response time for distinguishing neutral expressions from other emotions when viewing sketch faces. Our results showed that only the processing of neutral sketch faces was longer than for neutral photographed faces, whereas for the other emotions no differences regarding the reaction were observed for both stimuli. This finding is in line with studies that had observed shorter reaction times for the face identification of photographed faces in comparison to sketch faces when a neutral expression was used. (Meinhardt-Injac et al., 2013; Leder, 1996b; Zhao et al., 2016). The observed differences signalize that diagnostic features are important in transmitting the information of facial expressions when comparing sketch and photographed faces, but recognizing neutral expressions might require more information outside of the diagnostic feature regions.

Recent literature and investigations have led to the assumption that sketch faces reduce configural processing in comparison to their photographed counterparts (Zhao et al., 2016). Using sketch faces allow focusing on the sufficiency of the delivered information provided by the diagnostic features for face or emotion recognition. Presumably,

the main reason that the accuracy did not differ significantly between the stimulus types in our study can be attributed to the utilization of different face processing approaches (the visual scanning pattern will be described in the next section). In general, the accuracy ( $Hu$ ) is above 75% threshold across all facial expressions, demonstrating that the diagnostic features convey sufficient expressive information for emotion recognition regardless the presence of photographed facial texture information.

## 4.2 Fixation patterns

### 4.2.1 The mean fixation duration and location

One of the important contributions of this thesis to the research field of facial recognition is the first known investigation of the impact of configural information on facial expressions in a direct comparison between sketch and photographed face stimuli. Our results showed that participants located their fixations on the diagnostic features regardless the stimulus type during the entire face presentation. These findings demonstrate that the facial features provide salient cues for accurate expression recognition, even when facial texture and shape information are removed like for the sketch stimuli. Interestingly, the mean fixation duration and location showed no large differences between sketch and photographed faces, indicating that the extraction of information from the diagnostic features led to the equivalent performance in emotion recognition.

The distribution of the mean fixation revealed that different fixation patterns are applied for decoding specific facial expression. Regarding the fixation duration for the recognition of each emotion, fearful and angry expressions have led to longer duration in comparison to the happy expression. Thereby the location of fixations was mostly in the center of the face. Similar findings were also observed in a different study in which participants spent longer fixations on the nose region when recognizing fearful and angry expression ([Schurgin et al., 2014](#); [Bombari et al., 2013](#)). Additionally, participants

tend to locate their fixation more on the eyes and nose region when processing negative expressions, such as angry, fearful and sad expressions in this thesis. Although the smiling mouth of happy faces is a unique facial feature, we did not observe a longer fixation duration on this specific feature in comparison to the other facial features. However, on the other hand, the reaction time to recognize this emotion was rather short and thus led to a short mean fixation duration in comparison to other emotion, which complicates the analysis of the attention towards single features.

### **4.2.2 The first and second fixation duration and location**

To gain more information about the employed visual scanning strategy for each emotion, we examined the dynamic temporal organization of first and second fixations. Our results showed that the first fixation exhibits a tendency to be located on highly distinctive facial features, such as the smiling mouth on the happy faces. We observed that the smiling mouth captured the immediate fixation, whereas the fixation was evenly distributed on the center of the face for the other emotions. The extraction of information from a distinctive feature such as the mouth for happy faces is connected to part-based processing. Many studies show that looking for distinctive facial features leads to a better performance ([Calvo and Lundqvist, 2008](#)). The location of the first fixation for the other emotions indicates the utilization of configural processing. Important to note is that the first fixation duration and location were unaffected by the stimulus type. Thus reflecting that those stimuli share the same distinctive diagnostic feature (e.g., smiling mouth).

Regarding the relationship between emotions and facial features, the first fixation duration was long for the mouth on the happy expressions as for the mouth on the fearful faces. These results are in accordance with the study of [Eisenbarth and Alpers \(2011\)](#) in which they showed that the open mouth on fearful and happy faces is attracting the visual attention from participants in order to identify the unique facial feature at the very first place. However, participants looked not only at the mouth but also fixated on the nose region when recognizing fearful faces in the first fixation.

This fixation distribution on the two facial features exclude a preferred part-based processing of fearful expressions and shows that only for the happy expression pronounced part-based processing can be observed. The other facial expressions did not show any preferential facial features, in turn, a configural processing for examining the global information takes place. Conclusively and similar to what can be found in the literature, the first fixations show a systematic pattern of part-based processing only when a distinctive feature capturing the immediate visual attention ([Vaidya et al., 2014](#)). Otherwise, configural processing remains for further processing (e.g., further fixation inspection).

Concerning the second fixation, we found that the fixation patterns varied not only with emotions but also with stimulus type. The prevalence of the eyes was significant for the sketch face stimuli expressing anger, happiness, and sadness, whereas participants distributed their fixations evenly across diagnostic features on photographed faces across emotions. The strong influence of the eye region for recognizing sadness, happiness, and anger during the second fixation was also found in another study using high-spatial frequency face stimuli on healthy subjects ([Adolphs et al., 2005](#)). The newly created sketch faces might share similar characteristics as high-spatial frequency faces, like emphasizing the edge and contour of the facial features with rich details. Therefore, participants have orientated their visual attention to the eye region when processing sketch faces. Focusing on the extraction of information from the eye area might enhance the performance of impoverished facial stimuli and complex facial expressions. [Henderson et al. \(2005\)](#) suggests that participants exhibit a strong tendency to look at the eyes when they want to improve the face learning. Sketch faces are impoverished face stimuli; the eyes might provide more expressive information than the mouth and nose.

### 4.2.3 Summary

The results of the eye-tracking study demonstrate the importance of diagnostic features' presence. Because sketch faces contain the eyes, nose, and mouth as diagnostic features,

they transmit just as much information to recognize emotional expressions as the photographed face counterparts. Thereby the mean fixation locations of the participants showed a similar fixation on the diagnostic features independent of the utilized facial stimuli. In general, the first fixation is attracted by the facial expression carrying highly distinctive facial features (i.e., happy mouth) leading to part-based processing. In the absence of such a feature configural processing is employed. Differences between sketch and photographed stimuli were found in the analysis of the second fixation, which revealed that photographed faces exhibit still configural processing, whereas the part-based processing is more dominant on sketch faces. This is evidenced by retrieving information dominantly from the diagnostic features such as the eye region at the second fixation and confirms the assumption that sketch faces reducing the employment of configural processing. In sum, the results and analysis of the data obtained by eye-tracking show the importance of configural processing and part-based processing for emotion recognition.

### **4.3 Influence of sketch and photographed face stimuli on ERPs**

The ERP components P100 and N170 reflect time-locked visual information from the neural activity, yielding an accurate estimate of stimulus-locked response over cortical areas. Based on our first experiments, which have shown that the configural information is reduced on sketch faces, we aimed to further investigate the impact of the configural information on the sketch face by testing two experimental paradigms: face inversion and face deconfiguration. The evaluation of the effects of the two experimental paradigms on the P100 and N170 components can generate valuable insights: First, the P100 components are sensitive to the low visual properties and face configural processing. Second, the N170 amplitude is increased and delayed when the configural processing is interrupted.

Evaluating the inversion effect on the sketch stimuli has shown that N170

amplitudes did not increase. This finding suggests that sketch faces might activate different neural networks due to the difficulty to identify an emotional expression out of the sketch gestalt (i.e., physiognomic information when the face is inverted). Despite the fact that the typical face inversion effect was not observed on sketch faces, the N170 latency showed an additive effect, in which the brain treated sketch and photographed faces as two different categories.

Regarding the effect of face deconfiguration, it seems that configural processing is not relevant for emotion recognition on sketch faces as the N170 component is not affected. Furthermore, the accuracy of behavior performance did not varied either with different deconfiguration conditions. In the following, these main findings, implications, and the perspective for future works are discussed.

#### 4.3.1 P100.

The examination of P100 indicates that sketch faces are processed at the early stage of face processing like photographed faces but with some differences. We observed that P100 amplitudes were larger for photographed faces than for sketch faces. The inversion of the face stimuli has led to larger amplitudes for the inverted stimuli in comparison to the upright faces. However, no interaction between the face stimuli was observed. These findings suggest that inverted and photographed faces required additional neural responses to process the information of the face configuration. The implication of the P100 component as an indicator for configural processing is still subject to debate as it also could be only related to low-level visual differences (e.g., [Rossion and Boremanse 2008](#); [Rousselet et al. 2008](#)). Some studies suggest that the P100 is related to the rapid global processing of facial stimuli. [Batty and Taylor \(2003\)](#) found that the P100 did not show any differences between different facial expressions. But inverted and atypical face stimuli elicited a larger and delayed P100 in comparison to upright and typical face stimuli ([Itier and Taylor, 2004a](#); [Halit et al., 2000](#); [Herrmann et al., 2005](#)). Conclusively, P100 might be sensitive to the configuration of the face. Thus, early face processing in the P100 time window examines the relevant facial

components that locate in regions that do not violate the facial structure.

Our results of the face deconfiguration study also provide evidence that the spatial relationship between facial features is less influential for the P100. Despite the manipulation of the stimuli with six different degrees of face deconfiguration, the P100 component did not modulate significantly by the different form of configuration. The efficient extraction of configural information might take place at early face processing during the P100 time window. These findings support that the removal of facial texture reduces configural processing on sketch faces. To recognize faces efficiently, the prerequisite of configural processing requires an early, fast and efficient categorization of faces.

### 4.3.2 N170

A previous study was investigated photographed and sketched faces regarding the face inversion effect on the N170 component, however, only neutral facial expressions were used in their study ([Sagiv and Bentin, 2001](#)). We were interested how the presence of different emotions influences the N170 component. As mentioned in the discussion section in the *Publication 2*, the task in the study of [Sagiv and Bentin \(2001\)](#) was to count mentally presented objects, i.e., flowers, while passively viewing photographed and sketch faces in upright or inverted orientation. Contrary to this paradigm, our study aims to decipher the complex face perception strategy, and therefore we used an emotion recognition task, which increases the salience of facial features. In comparison to the former study, we obtained similar results regarding the behavior of the N170 amplitude in the presence of upright or inverted sketch faces across all emotions. Taken together, these findings illustrate that the N170 is robust to face stimuli and can be used as an indicator of different tasks (active or passive face recognition) for the presence of configural processing.

As expected we observed larger N170 amplitudes for inverted photographed faces in comparison to upright faces across the emotions. Interestingly, the N170

amplitude of inverted photographed faces was similar to the amplitude measured when participants recognized upright sketch faces. This is already the first indication that by removing the textural features the configural information of the sketch stimulus was degraded and reduces configural processing. However, the reaction time for recognizing an emotion was similar between photographed and sketch faces as well as scoring a high accuracy. These results demonstrate that different face processing strategies are required between photographed and sketch faces in order to decode efficiently facial expressions.

These findings are in line with the result of early studies for face processing in passive viewing situations by [Sagiv and Bentin \(2001\)](#) as well as the work by [Caharel et al. \(2013\)](#). Whereas all of these authors limited the task to less attention driven facial features, the good behavioral performance of the emotion recognition made it possible to reproduce the N170 effect and correlate it to the facial expression process in the present experiment. Furthermore, we could control the metric distance of facial features that can indirectly produce a dynamic facial expression or distort configuration.

Sketch faces generated a similar behavioral performance to photographed faces when presented in the upright orientation, which was not the case when recognizing facial expressions from inverted sketch faces. The data of our study show that participants are slower and less accurate in recognizing inverted sketch faces and brain recording ERPs showed a reverse N170 response between sketch and photographed faces. Furthermore, we observed a linear trend of reaction time for the sketch with different deconfiguration conditions but not for the N170 amplitude and latency. These results provide a first indication that the N170 is devoted to the configurational analysis of whole faces, and the N170 is elicited maximally by facial stimuli that are in optimal canonical view for recognition. Other types of face stimuli trigger the attenuation of N170 as a function of the degree to which these facial stimuli vary from the 'optimal face', particularly when presented in upright orientation ([Eimer, 2000a](#)). Thus, the N170 should also reflect the processing of the diagnostic internal features in terms of the configurational analysis of the whole face in order to be utilized by subsequent face processes (e.g., identification) ([Bruce and Young, 1986](#)).

## Relevance to event-related EEG studies.

By comparing sketch and photographed faces, the seemingly effortless emotion recognition resulted from different strategies for face processing. Interesting to note that upright sketch faces produced similar N170 amplitudes to inverted photographed faces, suggesting upright sketch faces depend more on part-based processing than on configural processing. The eye-tracking study in this thesis supported similar findings, the part-based processing took place during the second fixation on the eye region for sad, angry, and happy expressions. Future studies should investigate whether fixating on the eye region on the sketch faces is the main cause leading to larger N170 amplitudes. Some ERP studies demonstrated that the N170 is increased when internal facial features are presented, particularly when only the eyes are presented (Itier et al., 2006; Itier and Taylor, 2004a). The eyes appear to convey influential expressive information leading to a larger N170 and would explain the observation of a larger N170 for upright sketch faces. The eye region appears to be important in face processing because the eyes positioned in the upper part of the face where they might define a global frame of a face (Bentin et al., 2006). Therefore, the N170 can be influenced by a single diagnostic facial feature like the eyes (Itier et al., 2007, 2006; Eimer, 1998).

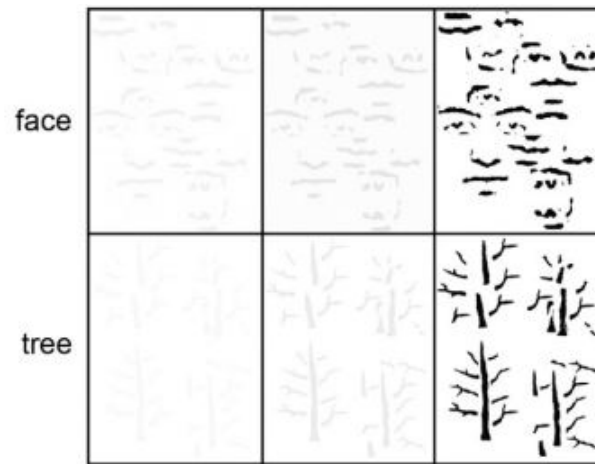
Many studies have shown that the eye region is one of the key areas where expressive information are extracted (Scheller et al., 2012; Whalen, 2004; Schurgin et al., 2014). More recent studies found that facial expressions engage different parts of the spatiotemporal network, for example, happy expression starts between 100-120 ms at occipital region whereas fearful expression begins around 150 ms at posterior-lateral sites (Neath-Tavares and Itier, 2016; Hinojosa et al., 2015; Itier and Neath-Tavares, 2017). They also found that the N170 is increased when participants fixated on the eye region regardless of the facial expression. The presence of eyes is more likely linked to a representation as a face configuration when the face is presented in a natural canonical orientation. Thus, the larger N170 activated by the eyes (Itier et al., 2006; Nemrodov et al., 2014; Bentin et al., 2006) might suggest the use of the eyes as an important cue for processing emotional sketch faces. Although we found that part-based processing occurred during the second fixation, we need further investigation to ensure it is

the main cause for part-based processing during the N170 time window for sketch faces. In this case, one might expect the difference in the onset latency of ERP and eye movement (e.g., the fixation and saccade) for emotional expressions elicited by different facial expression and different face stimuli.

## 4.4 Outlook and perspective

The new created set of sketch face stimuli has shown in the different experiments to convey the key information for emotion recognition. An advantage of this stimuli set is the easy manipulation of the diagnostic facial features, as demonstrated in the second EEG (i.e., deconfiguration paradigm), using systematic algorithms to avoid any artificial factor of manipulation or modification.

**Perspectives for sketch faces database** The created sketch faces database used here is highly suitable for anyone who is interested in manipulating diagnostic facial features (i.e., each feature is an individual segment). For example, in a very recent study, [Maher et al. \(2016\)](#) used sketch faces between trees to identify the lack of face-sensitive temporal response of the brain in patients with schizophrenia (see in the Figure 4.1). In this study, the face stimulus is very similar to our sketch face stimulus, but they neither used more than one sketch face stimulus nor emotional faces (i.e., only one neutral face stimulus). Besides, they lacked an instruction of how this face stimulus was generated, which is important for a better comparison. There are many natural photographed face databases as mentioned in Chapter 2.1, but less classified non-natural face stimuli datasets. A comparison between different non-photographed face stimuli has led to different conclusions on emotional face recognition. For example, angry superiority was found in line-drawing faces but not in photographed faces ([Horstmann and Bauland, 2006](#); [Elsherif et al., 2017](#)). The eye-brows on line-drawing faces create proximity of "V" shape as they are presenting as two diagonal lines, resulting in an advantage for participants to identify anger quicker. In sum, shows more research is necessary to draw further conclusions. The



**Figure 4.1:** The image illustrates three different level of stimulus contrast based on individual's psychophysical testing. From left: perceptual threshold, in the middle: two times perceptual threshold, and on the right: 100% contrast. The tasks were face detection and tree detection measured by EEG. Reprinted from [Maher et al. \(2016\)](#), *Schizophrenia Research*, with permission

here created database would be a good starting point for these investigations as it allows the direct comparison to the photographed original faces and also a simple manipulation of the diagnostic features.

#### **Assessment of emotion recognition with individuals with schizophrenia or autism.**

The inability to recognize facial emotions is the core feature of individuals with schizophrenia ([Bellgrove et al., 2003](#); [Lee et al., 2010](#)). The ability to decode and discriminate facial expressions accurately and quickly is an essential component of social interactions. To accurately recognize facial expressions, individuals with schizophrenia require more information from the facial areas than healthy controls ([Morris et al., 2009](#)), especially for recognizing negative facial expressions, such as fear and disgust ([Kohler et al., 2003](#)). To identify a fearful expression, schizophrenia patients rely on the less informative mouth region, which is contrary to healthy subjects as they look at the eye region ([Lee et al., 2010](#)). Studies which used “non-emotional face processing” in individuals with schizophrenia suggested an underlying mechanism may be responsible for the abnormal face recognition, such as attention deficit or preference in processing featural information (part-based processing) ([Watson, 2013](#)). Meaning that individuals with schizophrenia have limited access to configural information. However, we have shown that sketch faces convey sufficient expressive

information like photographed faces. The part-based processing was more dominant on sketch faces, which fits to individuals with schizophrenia as they pay more attention to facial features to gather more information. However, it is still unclear whether part-based processing is the responsible visual exploration strategy for this atypical facial emotion processing. Using the emotional sketch faces, we might discover the underlying neuron mechanism.

In addition to sketch faces, studying the response to manipulated stimulus configurations (i.e., using inverted or scrambled sketch faces) is another alternative for understanding how patients with schizophrenia process facial emotions. Inverted face stimuli prevent the configural processing of encoded facial expressions, resulting in a feature by feature analysis. Scrambled faces illustrate that subtle changes of distance and location of face parts (e.g., eyes above the nose) and can disturb configural face processing (Schwartz et al., 2002). Schizophrenia patients judge the different distances between eyes of two side-by-side faces poorly, but performing similarly to healthy controls for detecting facial feature changes, such as the shape of the eye or mouth (Baudouin et al., 2008). These results suggested that patients with schizophrenia have a poor global processing strategy, which is known as featural processing bias or a fragmented sensory processing style (Butler et al., 2008; Watanabe et al., 2003). However, patients with schizophrenia are able to focus their attention on global or local levels of features when instructed (Granholm et al., 1999). The most efficient top-down attentional allocation strategy is impaired in patients with schizophrenia unless the salience of the stimulus content is highlighted by manipulation or explicit instructions.

Eye-tracking is a widely-used approach for investigating how patients with schizophrenia use visual information to decode facial expression. This approach provides a temporal and spatial measures of how visual attention is shared among facial expression stimuli. When identifying facial expressions, patients with schizophrenia tend to avoid informative regions like the eyes and mouth, which is linked to poorer and less accurate emotional recognition (Loughland et al., 2004). Healthy subjects process the eye region on fearful faces faster and earlier than other regions (Yang et al., 2007). Avoiding eyes appears to impair recognizing fear. This demonstrates

the importance of retrieving accurate information on the relevant facial region for recognizing particular emotions.

The obtained results in this thesis can be used for investigating the face processing with eye-tracking markers (See Publication 1) and EEG markers (See Publication 2) on patients who have a poor ability to recognize emotional information as compared to our control group from this thesis (total N= 90). These markers are of importance as sketch faces will provide useful facial expression cues for evaluating psychiatric disorders in terms of face processing. The facial components of our sketch stimuli are separable and will be easily superimposed into the background with similar textures, facilitating the design of many classical gestalt paradigms (embedded figures, contour matching, etc.). To exemplify the relevance of such stimulus manipulation, another study used embedded neutral sketch faces with sketch tree branches to examine whether patients with schizophrenia can detect the face using EEG (see 4.4 and [Maher et al., 2016, 2015](#)). They found that N170 amplitudes were larger for viewing faces than trees on control subjects, whereas no difference was found for patients with schizophrenia between viewing trees and faces. Thereby the utilization of sketch faces is beneficial for patients with schizophrenia or patients with autistic spectrum disorder because they are characterized by atypical part-based (local/featural) processing and configural processing ([Koldewyn et al., 2013](#); [Happé and Frith, 2006](#)). Furthermore, sketch faces can be treated as a face used in the bubble face paradigm (see [Schyns et al., 2002](#)), but with direct measures from participant's feedback from the combination of EEG and Eye-tracking rather than controlling the number of bubbles to reveal area. This perspective is laid out in the following: the sketch face is designed to address the role of diagnostic feature and reduced configural processing. This thesis has gathered fruitful behavioral and cognitive markers of behavioral accuracy, reaction time, fixations, P100 and N170 ERPs for future experiments.

**Co-register EEG and Eye-tracking.** In this thesis, sketch faces were used and compared to their photographed face counterparts in order to understand where do participants extract information when the configural information is reduced. Sketch faces appear to induce more part-based processing during the second fixation and can

also be observed in the N170 time window. Although the cognitive processing differs between sketch and photographed faces at early face processing stage, it is hard to compare two results when precise temporal events are not synchronized between EEG and Eye-tracking study. A recent study combining Eye-tracking and EEG found that emotion recognition interacts with fixation location using different facial expressions. For example, ERP components started as early as 100-120 ms for happy expressions (during the P100 time window), while the ERP response to fearful expressions started at around 150 ms or after the N170 component (Neath-Tavares and Itier, 2016). Other eye-tracking studies also discovered spontaneous saccades towards the eyes of fearful or the mouth of happy faces when the presentation is short as 150 ms (Scheller et al., 2012; Gamer et al., 2013). Moreover, they also found that the N170 is sensitive to the eye region. Note that in their study photographed face stimuli were used. However, similar results were found when showing isolated sketch eyes (Bentin et al., 2006). These results point out the need for monitoring gaze fixation in emotion studying with ERPs, as different effects reflect by emotions might be missed by simply measuring those common ERP markers for face processing. The advantage of combining eye-tracking and EEG is to provide a real-time, precise and spatial measure. Synchronizing signals of EEG and the Eye-tracker is to use the shared events (triggers or message) sent during the experiment will be used to evaluate the quality of synchronization. After this, the system should be able to detect saccades and fixations together with EEG events. The constraint of this methodology, as with the vast majority of EEG studies, is the requirement that the subject does not move their eyes. Here the difficulty is inherent in obtaining accurate eye position data, and related artifacts from the recorded data should be removed by employing a regression-based EOG (electrooculogram) artifact reduction method (Schlögl et al., 2007). This methodology should allow the acquisition of clean EEG signals and eye position data under natural, overt viewing conditions. Therefore, this approach allows us to measure preparatory electrocortical activity (as measured by saccade-related potentials) between saccades directed toward to different diagnostic features (e.g., the fearful eyes or the mouth of the happy face) as compared to non-diagnostic facial features. Furthermore, this approach also allows us to measure fixation related potentials such as the electrocortical activity.

## 4.5 Conclusion

The utilization of sketch faces demonstrated that diagnostic features are crucial for processing facial expressions. The correct categorization of an expression relies on specific facial features but also on the integration of facial information. Furthermore, the experiments in this thesis show how the nature of different stimulus information (sketch and photographed), their orientation (inverted, upright) and their deconfiguration can provide a window into how the perception of facial expression is functioned. We have used sketch faces to report three things: (i) input to the visual system reflect a need to extract information of diagnostic facial features (eyes, nose, and mouth) to achieve emotion recognition correctly. This could depend on both configural and part-based processing. The results also show that sketch faces provide sufficient expressive information for recognizing fear, neutral, anger, sad and happy faces even with removed facial information and texture; (ii) detection of highly distinctive feature (i.e., smiling mouth) is crucial as to gather a global configural information to guide behavioral response rapidly and accurately. When configural information is less available (sketch faces), the featural information was more likely drawn from the eye regions; (iii) the orientation influences the emotion recognition performance more for sketch faces than for photographed faces. Upright sketch faces result in similar increased N170 amplitudes as inverted photographed faces, suggesting the configural processing is disrupted for processing upright sketch faces. However, the behavioral performance was the same between upright sketch and photographed faces. Despite the configuration of sketch faces were distorted, the accuracy was high, but RTs increased linearly. We have inferred the part-based processing of the N170 over occipitotemporal regions and related to behavioral judgments on sketch faces. An interaction of face stimuli information and orientation suggests that the processing of facial expressions benefits from featural information. Moreover, this extends to early configural processing, possible to integrate the complex visual information into higher-order face processing of stimulus features, in order to facilitate the optimal and efficient extraction of expressive content on the faces by the brain.



## Supplementary data

### 5.1 EEG pilot study on individuals with schizophrenia

#### 5.1.1 Hypotheses

Individuals with schizophrenia recognize emotions from facial features less well than healthy subjects, and this impairs their ability to interact with others. Patients with schizophrenia show ‘restricted’ scanning and reduced attention to diagnostic facial features such as eyes, nose, mouth. ([Bortolon et al., 2015](#); [Watson, 2013](#); [Lee et al., 2010](#); [Schwartz et al., 2002](#)). Studies have shown that individuals with schizophrenia tend to extract facial information from less informative parts of the face (i.e., part-based processing), whereas healthy subjects rely more on the global facial information (i.e., configural/holistic processing) ([Morris et al., 2009](#); [Granholm et al., 1999](#); [Maher et al., 2015](#)). More importantly, patients with schizophrenia are highly sensitive to certain negative facial expression, such as fear and anger ([Mandal et al., 1998](#); [Fakra et al., 2015](#)). These observations highlight the abnormalities in the early face processing and reflect a failure to integrate diagnostic features due to dysfunctions in the part-based processing of details ([Loughland et al., 2002](#)).

From the previous studies in this thesis, the selected sketch faces convey sufficient expression information for emotion recognition like their photographed face

counterparts. We demonstrated that sketch faces have a reduced configural information, resulting in the enhancement of the analysis of featural information. In this study, we want to apply the newly created sketch faces and their photographed face counterparts on individuals with schizophrenia. We assume that individuals with schizophrenia would recognize the emotional content by using the part-based information on sketch faces, which might facilitate the emotion recognition in comparison to the photographed faces. We also expect to observe a reduced activity of configural processing in comparison to our healthy controls. This pilot study followed the same experimental design as the EEG studies in the Publication 2 (see Chapter 3.5) to investigate if patients with schizophrenia may exhibit the expected behavior known from the literature and how it may also differ using the new created emotional sketch stimuli

### 5.1.2 Method

**Participants** Two male patients ( $M = 46.5$ ) with schizophrenia disorder were recruited and met the DSM- V-TR criteria for this study. Participants were recruited from the *Centre Hospitalier de Versailles*. The patients with schizophrenia in this study have not met the criteria for drug abuse, co-morbid neurological disorder, history of brain trauma. The control group had no history of psychosis or affective disorder diagnosis (MINI) and had a fluent level of French (fNART). Participants had a semi-structured clinical interview about: the estimation of premorbid intelligence levels in French speakers (fNART); positive and negative syndrome scale (PANSS); MINI international neuropsychiatric interview (M.I.N.I), and questionnaires of Edinburgh Handedness Inventory, the Calgary Depression Scale. The experiment was run according to the Declaration of Helsinki ethical guidelines. All participants had the normal or corrected-to-normal vision and signed informed consent to take part in the study.

**Experimental protocol and material.** Three sessions are contained in this study: 1) practice session; 2) face inversion session (same experimental design, see Chapter 3.4.2); 3) deconfiguration face session ( same experimental design see Chapter 3.4.3). The

task was a self-paced forced-choice emotion recognition task. Face stimuli used in the practice session were not used in the experimental sessions. The face inversion session contained 150 sketch or photographed face stimuli: 10 face stimuli (five upright, five inverted)  $\times$  five emotions  $\times$  three random repetitions. Deconfiguration face session consisted of 150 upright sketch face stimuli: 5 face identity  $\times$  5 emotions  $\times$  6 different deconfiguration pixel level (0, 5, 10, 15, 20, and 25). Block order was counterbalanced across participants. Each block was separated by a self-paced pause. Each trial began with a 600 ms fixation cross, followed by the facial stimulus after a 300 ms blank with white background. The facial stimulus stayed on the screen until the participant responded without feedback. Although the response time was self-paced with no time limit applied, participants were instructed to respond as quickly and accurately as possible and to avoid eye-blink during stimulus representation. All images were enclosed within a rectangular frame, in a 400  $\times$  400-pixel array (See Chapter 3.1.2 material for details). Each facial stimuli subtended a 10°  $\times$  10° visual angle displayed against a 1280  $\times$  1024 pixel white background and presented centrally on a 17-inch screen computer monitor (with a refresh rate of 60 Hz) using E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA). The head position to the display screen was at a 60 cm distance. The value of the visual angle resembles a normal face-to-face interaction (Henderson et al., 2005).

**EEG acquisition.** EEG was recorded using BrainAmp amplifier and Brain Recorder software (Brain Products, Munich, Germany) with 32 Ag/AgCl active scalp electrodes (Acticap). Electrodes were referenced to a common ground situated at the vertex. The eye movement was monitored by VEOG (vertical electrooculogram) with FP1 and FP2. Electrode impedance were maintained below 10k. The EEG signals were measured at a sampling rate of 1000 Hz and were filtered through a 50-Hz notch filter. Data were bandpass filtered offline between 0.1 and 40 Hz (24 dB per octave). To eliminate electrical artifacts, epochs with voltages exceeding  $\pm 50$   $\mu$ V at any electrode were excluded from analysis. Ocular artifacts were removed using ocular correction ICA after examining data for artifacts and marker respective time period using raw data inspection. The resulting average number of trials retained per condition 6 across 2 participants. Epochs were taken from 100 ms before visual stimuli onset to 1000 ms

after. The electrical baseline of an electrode was defined as the average voltage during the -100 to 0 segment.

### 5.1.3 Analyses

**Data Analysis** Behavioral data were analyzed regarding the RT and  $Hu$  as in the Publication 2 (see 3.5). Only correct trials were analyzed for behavioral data and ERPs data. In this pilot study, the sample size is relative small compared with the previous studies on healthy participants. In order to conduct ANOVA as in Publication 2, the Shaprio-Wilk test of distribution of normality were used. The results showed that both RT and  $Hu$  were are significantly different from normal distribution (RT:  $W = .90$ ,  $p < .0003$ ;  $Hu$ :  $W = .92$ ,  $p < .003$ ). Therefore, only the mean with standard deviation (SD) were provided for behavioral data. ERP data were analyzed as explained in Publication 2 in order to compare between the current results with the results of healthy subjects. However, only the grand average of ERPs data is provided for data visualization due to insufficient number ERPs trials after ICA corrections, raw data inspection, and only the corrected trials were analyzed (minimum three trials per condition in order to run grand average for ERP, participant 1 had the average trial  $n \leq 5$ , and participant 2 had the average trial  $n < 3$ ).

### 5.1.4 Summary of results

#### Behavioral results per stimuli.

The calculated unbiased hit rate ( $Hu$ ) showed an average of  $.62 \pm .29$  across the conditions and emotions. Subjects obtained higher  $Hu$  for upright stimuli ( $.76 \pm .20$ ) than for inverted stimuli ( $.44 \pm .38$ ), and higher  $Hu$  for photographed faces ( $.66 \pm .34$ ) than than for sketch faces ( $.54 \pm .31$ ) (see Table 5.1). The  $Hu$  values for deconfiguration

**Table 5.1:** Summary of RT and  $Hu$  per condition.

condition	$Hu$	$sd$	RT (ms)	$sd$
Upright photograph	.84	.17	2852	1153
Inverted photograph	.48	.37	3865	2058
Upright sketch	.67	.22	2875	1393
Inverted sketch	.39	.34	2810	1217
Deconfiguration sketch	.67	.17	2878	632

faces were close to upright sketch faces ( $.67 \pm .17$ ).

The reaction time (RT) was in average  $3007 \pm 1365$  ms across conditions. Subjects spent longer times in processing the inverted stimuli ( $3365 \pm 1752$  ms) than the upright stimuli ( $2864 \pm 1244$  ms). In general, subjects spent longer time on processing inverted photographed faces than the other conditions (upright sketch and photograph faces, inverted sketch faces, and deconfiguration faces), see Table 5.1.

### Behavioral results: Stimulus type x Emotions.

The recognition of happiness received the highest  $Hu$  ( $.94 \pm .003$ ) and anger had the lowest  $Hu$  ( $.28 \pm .41$ ) for upright photographs across emotions. For inverted photographed faces, happiness obtained highest  $Hu$  and neutral faces received the lowest  $Hu$  than other emotions, see Table 5.2. With regard to sketch faces, sad faces were recognized the best for both upright ( $.86 \pm .26$ ) and inverted ( $.83 \pm .26$ ) conditions. Neutral upright sketch faces received the lowest score of  $Hu$  ( $.50 \pm .33$ ), which is quite close to upright sketch happy faces ( $.52 \pm .22$ ). The recognition of inverted neutral faces obtained the lowest  $Hu$  with  $.17 \pm .23$ . The results in the deconfiguration condition show that the value of  $Hu$  was highest for recognizing happy faces, whereas lowest for recognizing sad faces, see Table 5.2.

Regarding the measured reaction times, subjects spent more time in responding to angry faces ( $3861 \pm 1477$ ), but less time for happy faces ( $1787 \pm 439$ ) than other emotions using the upright photographed face stimuli, see Table 5.2. For inverted

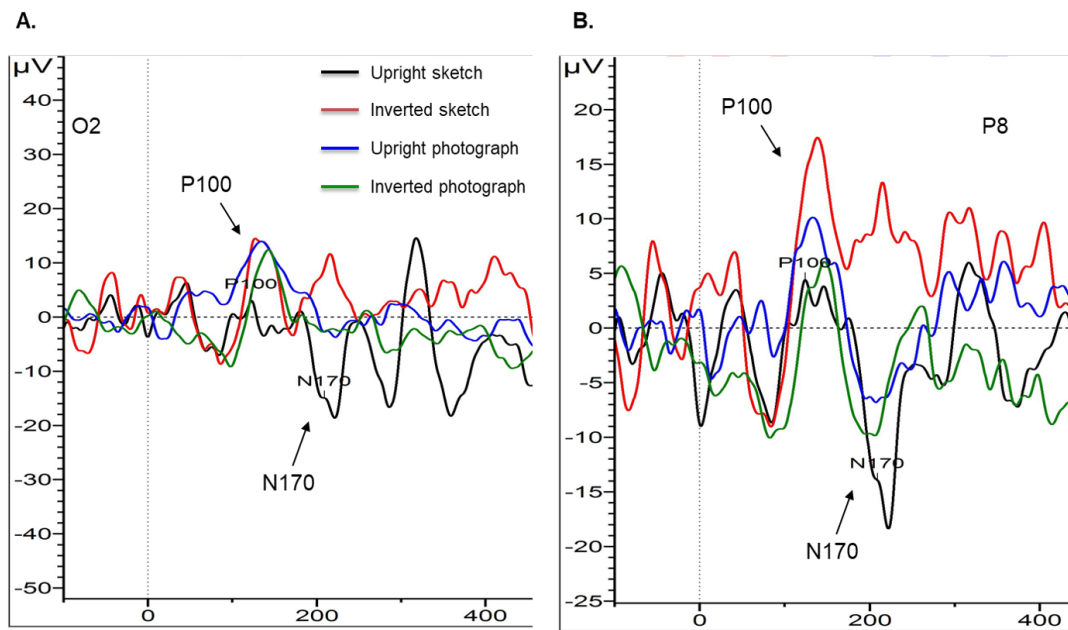
**Table 5.2:** Summary of RT and *Hu* with SD per Emotion x Condition

Condition	Angry	Fear	Happy	Neutral	Sad
<i>Hu</i>					
Upright photograph	.75 ± .19	.80 ± .19	.94 ± .003	.89 ± .16	.84 ± .23
Inverted photograph	.38 ± .41	.48 ± .48	.78 ± .22	.25 ± .05	.52 ± .68
Upright sketch	.71 ± .24	.78 ± .04	.52 ± .22	.50 ± .33	.86 ± .11
Inverted sketch	.09 ± .05	.40 ± .47	.33 ± .06	.17 ± .23	.83 ± .26
Deconfiguration sketch	.67 ± .27	.81 ± .24	.94 ± .09	.48 ± .26	.44 ± .29
RT					
Upright photograph	3861 ± 1477	2150 ± 656	1787 ± 439	3127 ± 1176	3338 ± 1364
Inverted photograph	3751 ± 1547	5314 ± 3360	2346 ± 1465	3053 ± 1808	4861 ± 2624
Upright sketch	2500 ± 455	3319 ± 1938	1450 ± 182	3331 ± 1632	3775 ± 2006
Inverted sketch	1947 ± 2754	3523 ± 1770	1958 ± 744	3099 ± 1512	2109 ± 1029
Deconfiguration sketch	3056 ± 759	2854 ± 820	1700 ± 436	2815 ± 905	2917 ± 1473

photographed face stimuli, the recognition of fearful faces resulted in the longest reaction times ( $5314 \pm 3360$ ), whereas the shortest reaction time was measured for categorizing happy faces ( $2346 \pm 1465$ ), see Table 5.2. Regarding sketch faces, the longest reaction times were observed in recognizing sad faces in upright orientation ( $3775 \pm 2006$ ), whereas recognizing fearful faces in inverted orientation ( $3523 \pm 1770$ ) was required the longest processing time. Participants responded faster for recognizing happy sketch faces across emotions in upright orientation ( $1985 \pm 744$ ), and for angry face in inverted orientation ( $1947 \pm 2754$ ). In the deconfiguration condition, the recognition of the happy face processed the fastest ( $1770 \pm 436$ ) and the longest reaction time was found for processing angry faces ( $3056 \pm 759$ ).

### EEG data.

Due to insufficient trials for further analyzing after processing filter, ICA, data inspection and taking into account, only the corrected responded trials, the grand average ERPs of the P100 and N170 were provided for visualizing the results. There not enough correct ERP trials for conducting the grand average on the conditions of Stimulus type (sketch, photograph) x Orientation (upright, inverted) for the participant 2 (number of trials:  $n < 3$ ). However, there was a minimum average of three trials for the deconfiguration condition. For participant 1, the valid trial number was average 10 across all the conditions. Therefore, only the data of the participant 1 were computed for grand

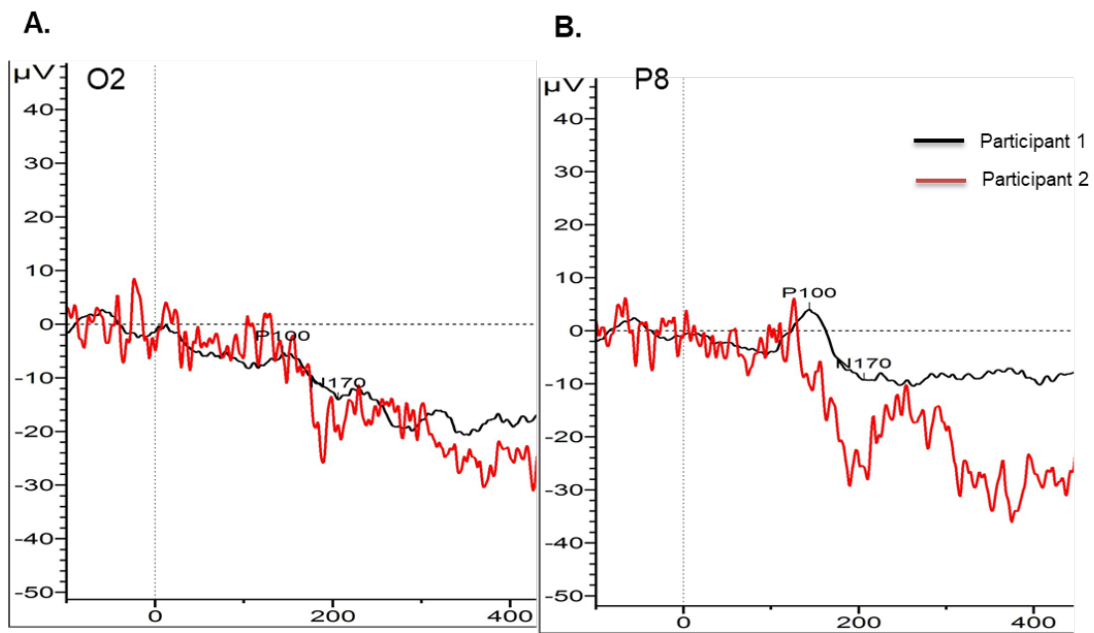


**Figure 5.1:** On left panel: P100 components at O2 channel; on the right panel: N170 components at P8 channel across conditions from participant 1.

average on Stimulus type (sketch, photograph)  $\times$  Orientation (upright, inverted), see Figure 5.1, and the data of both participants were computed for the deconfiguration condition, see Figure 5.2.

Visually inspection of the results for the P100 on Stimulus type (sketch, photograph)  $\times$  Orientation (upright, inverted) for the P100 (see Figure 5.1 A) showed that amplitudes were smaller for upright sketch faces than for other stimuli. During the N170 time-window, the upright sketch faces generated greater N170 amplitudes than other stimuli (see Figure 5.1 B). However, the N170 amplitudes were positive rather than negative for inverted sketch faces.

For the deconfiguration condition, the P100 and N170 components were not very clear, see Figure 5.2, particularly for participant 2.



**Figure 5.2:** On left panel: P100 components at O2 channel; on the right panel: N170 components at P8 channel across deconfiguration conditions from participant 1 (in black) & participant 2 (in red).

### 5.1.5 Conclusion and discussion

The purpose of this pilot study is to understand how patients with schizophrenia perceive emotions from sketch faces and photographed faces. When processing facial expressions on photographed face stimuli, happy expressions were recognized more accurate and rapid than other emotions for both inverted and upright orientation. Angry expressions were harder to be recognized on upright photographed faces, which have also been found for the healthy subjects in the previous Eye-tracking experiment (see Chapter 3.2). However, sad expressions were recognized better, and neutral expressions were poorly recognized on sketch faces regardless their orientation. Interestingly, when upright sketch faces were “deconfigural”, participants scored better for happiness and worse for sadness.

It is noteworthy that participants scored better for recognizing sad expressions on inverted sketch faces than inverted photographed faces, and required less time in processing the inverted sketch sad faces. Some studies have suggested that patients with schizophrenia were poor in the identification of sad expressions (Murphy and Cutting, 1990; Kosaka et al., 2002), and they did not show normal processing in the

amygdala activity during sad recognition in fMRI studies (Gur et al., 2002). Numerous studies have found that patients with schizophrenia have specific deficits in facial expression recognition, particularly in the processing of configural information, which is known as featural processing bias (Shin et al., 2007). The reduced configural information on sketch faces might induce different visual processing aiding the recognition of facial expressions, especially for sadness, by extracting more emotional information from the informative facial feature region. Inverted face stimuli prevent the configural processing of encoded facial expressions, resulting in a feature by feature analysis (Baudouin et al., 2008), and we found that patients with schizophrenia were recognizing well sad expressions when sketch faces were represented inverted. These findings suggest that the characteristics of sketch faces might engage different mechanisms for face processing in patients with schizophrenia. Especially, we observed that a largest N170 for processing the upright sketch faces than other face stimuli, but only inverted sketch faces were not generating N170 (produced a positive component). Note that, the N170 component is a negative component and reflects face configuration processing. Unfortunately, the results of the ERPs in deconfiguration condition were not very clear to draw any conclusion and to explain why the performances were better for recognizing happy expressions but worse for processing sadness on sketch deconfiguration.

The results for examining the effects of face processing on the perception of facial expressions were limited by the small number of participants with schizophrenia. Future studies should recruit more participants to compare with here obtained results. Also, future studies might reduce the number of stimuli per condition. In this study, we used 150 images per condition, and we found that the participant 2 fell asleep in-between response. Furthermore, it has been suggested that early sensory processing is associated with early dorsal visual stream processing, rather than ventral stream processing. (Foxe et al., 2001). The deficits in basic visual decoding in individuals with schizophrenia may be associated with poor working memory performance (Haenschel et al., 2007; Giersch et al., 2011). Unlike healthy subjects, we found participants exhibited a tendency to search for response keys each time before pressing the response key even they had passed the training session. Those constant movements of the head, hands, and eyes generated much noise, which resulted in insufficient ERP data for

analysis. Nevertheless, patients with schizophrenia are able to recognize the newly created sketch stimuli. We found that patients with schizophrenia have recognized well sad expressions on sketch faces, which is opposite to results in the literature. The largest N170 amplitude was generated for upright sketch faces and might suggest that participants benefit from the reduced configural information on sketch faces to recognize emotions. It will be interesting to further investigate how patients with schizophrenia process the part-based dominant sketch faces.

# Bibliography

- Adolphs, R. (2002). Neural systems for recognizing emotion. *Current opinion in neurobiology*, 12(2):169–177. [17](#)
- Adolphs, R. (2006). Perception and emotion: How we recognize facial expressions. *Current directions in psychological science*, 15(5):222–226. [27](#), [30](#)
- Adolphs, R., Baron-Cohen, S., and Tranel, D. (2002). Impaired recognition of social emotions following amygdala damage. *Journal of cognitive neuroscience*, 14(8):1264–1274. [12](#)
- Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., and Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, 433(7021):68–72. [21](#), [27](#), [67](#), [163](#), [168](#)
- Adrian, E. D. and Matthews, B. H. (1934). The berger rhythm: potential changes from the occipital lobes in man. *Brain*, 57(4):355–385. [38](#)
- Allport, F. H. and Allport, G. W. (1921). Personality traits: Their classification and measurement. *The Journal of Abnormal Psychology and Social Psychology*, 16(1):6. [8](#)
- Aragón, O. R. (2017). “tears of joy” and “tears and joy?” personal accounts of dimorphous and mixed expressions of emotion. *Motivation and Emotion*, pages 1–23. [7](#)
- Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., and Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology*, 13(4):379–398. [2](#)
- Barrett, L. F., Mesquita, B., and Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, 20(5):286–290. [8](#)
- Bartlett, J. C. and Searcy, J. (1993). Inversion and configuration of faces. *Cognitive psychology*, 25(3):281–316. [45](#)

- Bartlett, J. C., Searcy, J. H., and Abdi, H. (2003). What are the routes to face recognition. *Perception of faces, objects and scenes: Analytic and holistic processes*, pages 21–52. [17](#)
- Batty, M. and Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research*, 17(3):613–620. [4](#), [170](#)
- Baudouin, J.-Y., Vernet, M., and Franck, N. (2008). Second-order facial information processing in schizophrenia. *Neuropsychology*, 22(3):313. [46](#), [176](#), [189](#)
- Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., and Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. *Cognition & emotion*, 28(3):416–432. [23](#), [25](#)
- Beaupré, M. G. and Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of cross-cultural psychology*, 36(3):355–370. [59](#)
- Bediou, B., Franck, N., Saoud, M., Baudouin, J.-Y., Tiberghien, G., Daléry, J., and d’Amato, T. (2005). Effects of emotion and identity on facial affect processing in schizophrenia. *Psychiatry research*, 133(2):149–157. [2](#), [3](#), [21](#), [28](#)
- Behrmann, M. and Avidan, G. (2005). Congenital prosopagnosia: face-blind from birth. *Trends in cognitive sciences*, 9(4):180–187. [24](#)
- Bell, C. (1806). *Essays on the Anatomy of Expression in Painting*. Longman, Hurst, Rees, and Orme. [9](#)
- Bellgrove, M. A., Vance, A., and Bradshaw, J. L. (2003). Local–global processing in early-onset schizophrenia: Evidence for an impairment in shifting the spatial scale of attention. *Brain and cognition*, 51(1):48–65. [175](#)
- Benson, P. J. and Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, 3(1):105–135. [23](#)
- Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, 8(6):551–565. [3](#), [38](#), [41](#)
- Bentin, S. and Deouell, L. Y. (2000). Structural encoding and identification in face processing: Erp evidence for separate mechanisms. *Cognitive neuropsychology*, 17(1-3):35–55. [16](#), [41](#)
- Bentin, S., Golland, Y., Flevaris, A., Robertson, L. C., and Moscovitch, M. (2006). Processing the trees and the forest during initial stages of face perception: Electrophysiological evidence. *Journal of Cognitive Neuroscience*, 18(8):1406–1421. [19](#), [39](#), [40](#), [41](#), [45](#), [173](#), [178](#)

- Berger, H. (1929). Electroencephalogram in humans. *Archiv fur Psychiatrie und nervenkrankheiten*, 87:527–570. [38](#)
- Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., and Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal behavior*, 21(1):3–21. [59](#)
- Blais, C., Roy, C., Fiset, D., Arguin, M., and Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, 50(12):2830–2838. [9](#), [27](#)
- Blau, V. C., Maurer, U., Tottenham, N., and McCandliss, B. D. (2007). The face-specific n170 component is modulated by emotional facial expression. *Behavioral and brain functions*, 3(1):7. [4](#), [28](#)
- Boehm, S. G., Dering, B., and Thierry, G. (2011). Category-sensitivity in the n170 range: a question of topography and inversion, not one of amplitude. *Neuropsychologia*, 49(7):2082–2089. [44](#)
- Boll, S. and Gamer, M. (2016). Psychopathic traits affect the visual exploration of facial expressions. *Biological psychology*, 117:194–201. [12](#)
- Bombari, D., Mast, F. W., and Lobmaier, J. S. (2009). Featural, configural, and holistic face-processing strategies evoke different scan patterns. *Perception*, 38(10):1508–1521. [14](#), [32](#)
- Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., and Lobmaier, J. S. (2013). Emotion recognition: The role of featural and configural face information. *The Quarterly Journal of Experimental Psychology*, 66(12):2426–2442. [2](#), [3](#), [8](#), [14](#), [15](#), [16](#), [19](#), [20](#), [23](#), [25](#), [28](#), [31](#), [32](#), [33](#), [34](#), [63](#), [64](#), [66](#), [67](#), [164](#), [166](#)
- Bortolon, C., Capdevielle, D., and Raffard, S. (2015). Face recognition in schizophrenia disorder: A comprehensive review of behavioral, neuroimaging and neurophysiological studies. *Neuroscience & Biobehavioral Reviews*, 53:79–107. [12](#), [28](#), [39](#), [181](#)
- Bötzel, K., Schulze, S., and Stodieck, S. R. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Experimental brain research*, 104(1):135–143. [41](#)
- Boutsen, L., Humphreys, G. W., Praamstra, P., and Warbrick, T. (2006). Comparing neural correlates of configural processing in faces and objects: an erp study of the thatcher illusion. *Neuroimage*, 32(1):352–367. [39](#), [40](#), [44](#)
- Bruce, V. and Young, A. (1986). Understanding face recognition. *British journal of psychology*, 77(3):305–327. [20](#), [172](#)

- Bruce, V. and Young, A. W. (2012). *Face perception*. Psychology Press. 1, 16
- Buchan, J. N., Paré, M., and Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, 2(1):1–13. 32, 33
- Butler, P. D., Tambini, A., Yovel, G., Jalbrzikowski, M., Ziwich, R., Silipo, G., Kanwisher, N., and Javitt, D. C. (2008). What's in a face? effects of stimulus duration and inversion on face processing in schizophrenia. *Schizophrenia research*, 103(1):283–292. 176
- Caharel, S., Jiang, F., Blanz, V., and Rossion, B. (2009). Recognizing an individual face: 3d shape contributes earlier than 2d surface reflectance information. *Neuroimage*, 47(4):1809–1818. 4
- Caharel, S., Leleu, A., Bernard, C., Viggiano, M.-P., Lalonde, R., and Rebaï, M. (2013). Early holistic face-like processing of arcimboldo paintings in the right occipito-temporal cortex: evidence from the n170 erp component. *International Journal of Psychophysiology*, 90(2):157–164. 3, 16, 43, 172
- Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., and Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? evidence for a defect in representing diagnostic facial information following brain damage. *Journal of cognitive neuroscience*, 17(10):1652–1666. 12, 13
- Caldara, R., Zhou, X., and Miellet, S. (2010). Putting culture under the 'spotlight' reveals universal information use for face recognition. *PLoS One*, 5(3):e9708. 33
- Calder, A. J. and Jansen, J. (2005). Configural coding of facial expressions: The impact of inversion and photographic negative. *Visual Cognition*, 12(3):495–518. 17, 19
- Calder, A. J. and Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6(8):641–651. 67
- Calder, A. J., Young, A. W., Keane, J., and Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human perception and performance*, 26(2):527. 17, 18, 19, 20, 22, 23
- Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., and Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, 3(2):81–118. 12
- Calvo, M. G., Fernández-Martín, A., and Nummenmaa, L. (2014). Facial expression recognition in peripheral versus central vision: Role of the eyes and the mouth. *Psychological research*, 78(2):180–195. 28, 163

- Calvo, M. G. and Lundqvist, D. (2008). Facial expressions of emotion (kdef): Identification under different display-duration conditions. *Behavior research methods*, 40(1):109–115. [27](#), [36](#), [167](#)
- Calvo, M. G. and Marrero, H. (2009). Visual search of emotional faces: The role of affective content and featural distinctiveness. *Cognition and Emotion*, 23(4):782–806. [27](#)
- Calvo, M. G. and Nummenmaa, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *Journal of Experimental Psychology: General*, 137(3):471. [16](#), [163](#), [164](#)
- Calvo, M. G. and Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, 30(6):1081–1106. [9](#), [27](#)
- Calvo, M. G., Nummenmaa, L., and Avero, P. (2010). Recognition advantage of happy faces in extrafoveal vision: Featural and affective processing. *Visual Cognition*, 18(9):1274–1297. [22](#), [27](#), [30](#)
- Cangöz, B., Altun, A., Aşkar, P., Baran, Z., and Mazman, S. G. (2013). Examining the visual screening patterns of emotional facial expressions with gender, age and lateralization. *Journal of Eye Movement Research*, 6(4). [30](#)
- Carey, S., Diamond, R., et al. (1977). From piecemeal to configurational representation of faces. *Science*, 195(4275):312–314. [45](#)
- Carmel, D. and Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition*, 83(1):1–29. [43](#)
- Churches, O., Nicholls, M., Thiessen, M., Kohler, M., and Keage, H. (2014). Emoticons in mind: An event-related potential study. *Social neuroscience*, 9(2):196–202. [7](#), [8](#)
- Clark, C. M., Gosselin, F., and Goghari, V. M. (2013). Aberrant patterns of visual facial information usage in schizophrenia. *Journal of abnormal psychology*, 122(2):513. [2](#), [21](#)
- Cohn, J. F., Ambadar, Z., and Ekman, P. (2007). Observer-based measurement of facial expression with the facial action coding system. *The handbook of emotion elicitation and assessment*, pages 203–221. [9](#)
- Cole, H. W. and Ray, W. J. (1985). Eeg correlates of emotional tasks related to attentional demands. *International Journal of Psychophysiology*, 3(1):33–41. [38](#)
- Collishaw, S. M. and Hole, G. J. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception*, 31(3):287–296. [24](#)

- Curby, K. M., Johnson, K. J., and Tyson, A. (2012). Face to face with emotion: Holistic face processing is modulated by emotional state. *Cognition & emotion*, 26(1):93–102. [19](#)
- Darwin, C. (1998). *The expression of the emotions in man and animals*. Oxford University Press, USA. [9](#), [26](#)
- Darwin, C., Adler, M. J., and Hutchins, R. M. (2009). *The origin of species by means of natural selection*. Floating Press. [8](#)
- Davies, G., Ellis, H., and Shepherd, J. (1977). Cue saliency in faces as assessed by the 'photofit' technique. *Perception*, 6(3):263–269. [12](#)
- De Haan, M., Pascalis, O., and Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of cognitive neuroscience*, 14(2):199–209. [1](#)
- De Sonnevile, L., Verschoor, C., Njiokiktjien, C., Op het Veld, V., Toorenaar, N., and Vranken, M. (2002). Facial identity and facial emotions: speed, accuracy, and processing strategies in children and adults. *Journal of Clinical and experimental neuropsychology*, 24(2):200–213. [163](#)
- de Xivry, J.-J. O., Missal, M., and Lefèvre, P. (2008). A dynamic representation of target motion drives predictive smooth pursuit during target blanking. *Journal of Vision*, 8(15):6–6. [34](#), [35](#)
- Derntl, B., Seidel, E.-M., Kainz, E., and Carbon, C.-C. (2009). Recognition of emotional expressions is affected by inversion and presentation time. *Perception*, 38(12):1849–1862. [25](#)
- Diamond, R. and Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, 115(2):107. [17](#)
- Dien, J. (1998). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain topography*, 11(1):43–55. [117](#)
- Duchaine, B. and Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science*, 1:393–416. [19](#)
- Eimer, M. (1998). Does the face-specific n170 component reflect the activity of a specialized eye processor? *Neuroreport*, 9(13):2945–2948. [4](#), [44](#), [173](#)
- Eimer, M. (2000a). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical neurophysiology*, 111(4):694–705. [3](#), [41](#), [117](#), [172](#)

- Eimer, M. (2000b). The face-specific n170 component reflects late stages in the structural encoding of faces. *Neuroreport*, 11(10):2319–2324. [44](#)
- Eimer, M. and Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, 45(1):15–31. [3](#)
- Eisenbarth, H. and Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion*, 11(4):860. [3](#), [13](#), [14](#), [23](#), [26](#), [27](#), [29](#), [30](#), [32](#), [63](#), [66](#), [167](#)
- Ekman, P. (1972). Universal and cultural difference in facial expression of emotion. Nebraska symposium on motivation, Lincoln, NE: university of Nebraska Press. [8](#), [9](#)
- Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200. [9](#)
- Ekman, P. and Friesen, W. (1978). Facial action coding system: a technique for the measurement of facial movement. *Palo Alto: Consulting Psychologists*. [7](#), [9](#), [33](#)
- Ekman, P. and Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1):49–98. [14](#)
- Ekman, P. and Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1):56–75. [9](#), [10](#), [26](#)
- Ekman, P. and Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. *Ishk*. [9](#), [27](#)
- Ekman, P., Friesen, W. V., and Hager, J. C. (2002). *FACS investigator's guide. A human face*, page 96. [9](#)
- Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712. [16](#)
- Ekman, P., Friesen, W. V., and Tomkins, S. S. (1971). Facial affect scoring technique: A first validity study. *Semiotica*, 3(1):37–58. [8](#)
- Elfenbein, H. A. and Ambady, N. (2003). When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *Journal of personality and social psychology*, 85(2):276. [26](#), [59](#)
- Ellis, H. (1986). Processes underlying face recognition. *The neuropsychology of face perception and facial expression*. [12](#)

- Elsherif, M. M., Saban, M. I., and Rotshtein, P. (2017). The perceptual saliency of fearful eyes and smiles: A signal detection study. *PloS one*, 12(3):e0173199. [10](#), [12](#), [174](#)
- Eng, Z., Yick, Y., Guo, Y., Xu, H., Reiner, M., Cham, T., and Chen, S. (2017). 3d faces are recognized more accurately and faster than 2d faces, but with similar inversion effects. *Vision research*, 138:78–85. [25](#)
- Fakra, E., Jouve, E., Guillaume, F., Azorin, J.-M., and Blin, O. (2015). Relation between facial affect recognition and configural face processing in antipsychotic-free schizophrenia. *Neuropsychology*, 29(2):197. [181](#)
- Fiorentini, C. and Viviani, P. (2009). Perceiving facial expressions. *Visual Cognition*, 17(3):373–411. [3](#), [16](#)
- Flevaris, A. V., Robertson, L. C., and Bentin, S. (2008). Using spatial frequency scales for processing face features and face configuration: an erp analysis. *Brain research*, 1194:100–109. [19](#), [41](#), [45](#)
- Fox, E., Lester, V., Russo, R., Bowles, R., Pichler, A., and Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition & emotion*, 14(1):61–92. [27](#)
- Foxe, J. J., Doniger, G. M., and Javitt, D. C. (2001). Early visual processing deficits in schizophrenia: impaired p1 generation revealed by high-density electrical mapping. *Neuroreport*, 12(17):3815–3820. [189](#)
- Fraser, I. H., Craig, G. L., and Parker, D. M. (1990). Reaction time measures of feature saliency in schematic faces. *Perception*, 19(5):661–673. [19](#)
- Freire, A., Lee, K., and Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, 29(2):159–170. [24](#), [25](#)
- Frois-Wittman, J. (1930). The judgment of facial expression. *Journal of Experimental Psychology*, 13(2):113. [9](#)
- Gamer, M., Schmitz, A. K., Tittgemeyer, M., and Schilbach, L. (2013). The human amygdala drives reflexive orienting towards facial features. *Current Biology*, 23(20):R917–R918. [12](#), [178](#)
- Gauthier, I., Anderson, A., Skudlarski, P., and Gore, J. (2000). Expertise for cars and birds recruits right hemisphere face areas. *J. Cogn. Neurosci*, 19. [2](#), [44](#)
- Gendron, M. and Barrett, L. F. (2017). Facing the past. *The Science of Facial Expression*, page 15. [7](#), [41](#)

- George, N., Jemel, B., Fiori, N., Chaby, L., and Renault, B. (2005). Electrophysiological correlates of facial decision: insights from upright and upside-down mooney-face perception. *Cognitive Brain Research*, 24(3):663–673. [13](#), [41](#)
- Giersch, A., van Assche, M., Huron, C., and Luck, D. (2011). Visuo-perceptual organization and working memory in patients with schizophrenia. *Neuropsychologia*, 49(3):435–443. [21](#), [189](#)
- Goeleven, E., De Raedt, R., Leyman, L., and Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, 22(6):1094–1118. [26](#), [28](#), [53](#), [59](#)
- Goffaux, V., Gauthier, I., and Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research*, 16(3):416–424. [39](#), [43](#), [44](#)
- Goffaux, V. and Rossion, B. (2006). Faces are “spatial”—holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4):1023. [2](#), [16](#), [18](#), [22](#)
- Goren, D. and Wilson, H. R. (2006). Quantifying facial expression recognition across viewing conditions. *Vision Research*, 46(8-9):1253–1262. [63](#)
- Gosselin, F. and Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision research*, 41(17):2261–2271. [3](#), [7](#), [12](#)
- Grand, R. L., Mondloch, C. J., Maurer, D., and Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science*, 15(11):762–768. [22](#)
- Granholm, E., Perry, W., Filoteo, J. V., and Braff, D. (1999). Hemispheric and attentional contributions to perceptual organization deficits on the global-local task in schizophrenia. *Neuropsychology*, 13(2):271. [176](#), [181](#)
- Guo, K., Mahmoodi, S., Robertson, R. G., and Young, M. P. (2006). Longer fixation duration while viewing face images. *Experimental Brain Research*, 171(1):91–98. [1](#)
- Gur, R. E., McGrath, C., Chan, R. M., Schroeder, L., Turner, T., Turetsky, B. I., Kohler, C., Alsop, D., Maldjian, J., Ragland, J. D., et al. (2002). An fmri study of facial emotion processing in patients with schizophrenia. *American Journal of Psychiatry*, 159(12):1992–1999. [189](#)
- Haenschel, C., Bittner, R. A., Haertling, F., Rotarska-Jagiela, A., Maurer, K., Singer, W., and Linden, D. E. (2007). Contribution of impaired early-stage visual processing to working memory dysfunction in adolescents with schizophrenia: a study with

- event-related potentials and functional magnetic resonance imaging. *Archives of general psychiatry*, 64(11):1229–1240. [189](#)
- Halit, H., de Haan, M., and Johnson, M. H. (2000). Modulation of event-related potentials by prototypical and atypical faces. *Neuroreport*, 11(9):1871–1875. [39](#), [170](#)
- Hanawalt, N. G. (1944). The role of the upper and the lower parts of the face as a basis for judging facial expressions: Ii. in posed expressions and “candid-camera” pictures. *The Journal of General Psychology*, 31(1):23–36. [12](#)
- Handy, T. C. (2005). *Event-related potentials: A methods handbook*. MIT press. [38](#)
- Happé, F. and Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, 36(1):5–25. [177](#)
- Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6):223–233. [1](#), [20](#)
- Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological psychiatry*, 51(1):59–67. [16](#)
- Henderson, J. M., Williams, C. C., and Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & cognition*, 33(1):98–106. [11](#), [29](#), [56](#), [67](#), [168](#), [183](#)
- Henderson, R. M., McCulloch, D. L., and Herbert, A. M. (2003). Event-related potentials (erps) to schematic faces in adults and children. *International Journal of Psychophysiology*, 51(1):59–67. [16](#), [29](#), [40](#), [41](#), [45](#)
- Herrmann, M., Ehlis, A.-C., Ellgring, H., and Fallgatter, A. (2005). Early stages (p100) of face perception in humans as measured with event-related potentials (erps). *Journal of neural transmission*, 112(8):1073–1081. [39](#), [40](#), [41](#), [170](#)
- Hinojosa, J., Mercado, F., and Carretié, L. (2015). N170 sensitivity to facial expression: a meta-analysis. *Neuroscience & Biobehavioral Reviews*, 55:498–509. [24](#), [173](#)
- Hoffman, E. A. and Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature neuroscience*, 3(1):80. [20](#)
- Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, 23(1):65–74. [24](#)
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford. [29](#), [32](#)

- Horstmann, G. and Bauland, A. (2006). Search asymmetries with real faces: testing the anger-superiority effect. *Emotion*, 6(2):193. [27](#), [63](#), [174](#)
- Hsiao, J. H.-w. and Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, 19(10):998–1006. [19](#), [32](#), [33](#), [34](#), [56](#)
- Itier, R. J., Alain, C., Sedore, K., and McIntosh, A. R. (2007). Early face processing specificity: It's in the eyes! *Journal of Cognitive Neuroscience*, 19(11):1815–1826. [4](#), [13](#), [41](#), [44](#), [173](#)
- Itier, R. J., Latinus, M., and Taylor, M. J. (2006). Face, eye and object early processing: what is the face specificity? *Neuroimage*, 29(2):667–676. [45](#), [173](#)
- Itier, R. J. and Neath-Tavares, K. N. (2017). Effects of task demands on the early neural processing of fearful and happy facial expressions. *Brain Research*, 1663:38–50. [173](#)
- Itier, R. J. and Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using erps. *Neuroimage*, 15(2):353–372. [39](#), [40](#), [41](#)
- Itier, R. J. and Taylor, M. J. (2004a). Face recognition memory and configural processing: a developmental erp study using upright, inverted, and contrast-reversed faces. *Journal of cognitive neuroscience*, 16(3):487–502. [4](#), [39](#), [170](#), [173](#)
- Itier, R. J. and Taylor, M. J. (2004b). N170 or n1? spatiotemporal differences between object and face processing using erps. *Cerebral cortex*, 14(2):132–142. [41](#), [113](#)
- Itier, R. J., Van Roon, P., and Alain, C. (2011). Species sensitivity of early face and eye processing. *NeuroImage*, 54(1):705–713. [41](#)
- Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. [9](#)
- Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., and Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, 19(18):1543–1548. [3](#), [13](#), [14](#), [20](#), [30](#)
- Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., and Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19):7241–7244. [9](#)
- Jacques, C. and Rossion, B. (2006). Electrophysiological evidence for temporal dissociation between spatial attention and sensory competition during human face processing. *Cerebral Cortex*, 17(5):1055–1065. [7](#), [14](#)

- Jacques, C. and Rossion, B. (2009). The initial representation of individual faces in the right occipito-temporal cortex is holistic: electrophysiological evidence from the composite face illusion. *Journal of Vision*, 9(6):8–8. [41](#)
- Jasper, H. H. and Carmichael, L. (1935). Electrical potentials from the intact human brain. *Science*, 81(2089):51–53. [38](#)
- Jones, W., Carr, K., and Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of general psychiatry*, 65(8):946–954. [30](#)
- Kendall, L., Raffaelli, Q., Kingstone, A., and Todd, R. M. (2016). Iconic faces are not real faces: enhanced emotion detection and altered neural processing as faces become more iconic. *Cognitive Research: Principles and Implications*, 1(1):19. [41](#)
- Kimchi, R. and Amishav, R. (2010). Faces as perceptual wholes: The interplay between component and configural properties in face processing. *Visual Cognition*, 18(7):1034–1062. [25](#)
- Kobayashi, H. and Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, 387(6635):767. [13](#)
- Kobayashi, H. and Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of human evolution*, 40(5):419–435. [13](#)
- Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., Gur, R. E., and Gur, R. C. (2003). Facial emotion recognition in schizophrenia: intensity effects and error pattern. *American Journal of Psychiatry*, 160(10):1768–1774. [3](#), [175](#)
- Koldewyn, K., Jiang, Y. V., Weigelt, S., and Kanwisher, N. (2013). Global/local processing in autism: Not a disability, but a disinclination. *Journal of autism and developmental disorders*, 43(10):2329–2340. [177](#)
- Kosaka, H., Omori, M., Murata, T., Iidaka, T., Yamada, H., Okada, T., Takahashi, T., Sadato, N., Itoh, H., Yonekura, Y., et al. (2002). Differential amygdala response during facial recognition in patients with schizophrenia: an fmri study. *Schizophrenia Research*, 57(1):87–95. [188](#)
- Kowler, E. and Steinman, R. M. (1979). The effect of expectations on slow oculomotor control—i. periodic target steps. *Vision research*, 19(6):619–632. [29](#)
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., and Van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and emotion*, 24(8):1377–1388. [1](#), [4](#), [8](#), [9](#), [10](#), [21](#), [26](#), [28](#), [46](#), [53](#), [55](#), [59](#), [161](#), [164](#)

- Latinus, M., Love, S. A., Rossi, A., Parada, F. J., Huang, L., Conty, L., George, N., James, K., and Puce, A. (2015). Social decisions affect neural activity to perceived dynamic gaze. *Social cognitive and affective neuroscience*, 10(11):1557–1567. [13](#)
- Latinus, M. and Taylor, M. J. (2005). Holistic processing of faces: learning effects with mooney faces. *Journal of Cognitive Neuroscience*, 17(8):1316–1327. [3](#), [15](#), [40](#), [41](#), [43](#), [45](#)
- Latinus, M. and Taylor, M. J. (2006). Face processing stages: impact of difficulty and the separation of effects. *Brain research*, 1123(1):179–187. [25](#), [41](#), [43](#)
- Le Grand, R., Mondloch, C. J., Maurer, D., and Brent, H. P. (2001). Neuroperception: Early visual experience and face processing. *Nature*, 410(6831):890. [22](#), [23](#), [24](#)
- Leder, H. (1996a). Line drawings of faces reduce configural processing. *Perception*, 25(3):355–366. [17](#), [22](#)
- Leder, H. (1996b). Line drawings of faces reduce configural processing. *Perception*, 25(3):355–366. [165](#)
- Leder, H. and Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology: Section A*, 53(2):513–536. [1](#), [16](#), [23](#), [24](#)
- Lee, J., Gosselin, F., Wynn, J. K., and Green, M. F. (2010). How do schizophrenia patients use visual information to decode facial emotion? *Schizophrenia bulletin*, 37(5):1001–1008. [30](#), [46](#), [175](#), [181](#)
- Leppänen, J. M. and Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological research*, 69(1-2):22–29. [30](#), [59](#), [63](#)
- Leppänen, J. M., Milders, M., Bell, J., Terriere, E., and Hietanen, J. K. (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research*, 128(2):123–133. [27](#), [67](#)
- Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., and Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto-and electroencephalography. *Neuroscience letters*, 253(3):147–150. [39](#), [40](#)
- Liu, J., Harris, A., and Kanwisher, N. (2002). Stages of processing in face perception: an meg study. *Nature neuroscience*, 5(9):910–916. [39](#), [44](#)
- Liu-Shuang, J., Ales, J., Rossion, B., and Norcia, A. M. (2015). Separable effects of inversion and contrast-reversal on face detection thresholds and response functions: A sweep vep study. *Journal of vision*, 15(2):11–11. [45](#)

- Liversedge, S. P., Paterson, K. B., and Pickering, M. J. (1998). Eye movements and measures of reading time. *Eye guidance in reading and scene perception*, pages 55–75. [23](#), [29](#)
- Loughland, C. M., Williams, L. M., and Gordon, E. (2002). Visual scanpaths to positive and negative facial emotions in an outpatient schizophrenia sample. *Schizophrenia research*, 55(1):159–170. [181](#)
- Loughland, C. M., Williams, L. M., and Harris, A. W. (2004). Visual scanpath dysfunction in first-degree relatives of schizophrenia probands: evidence for a vulnerability marker? *Schizophrenia research*, 67(1):11–21. [3](#), [176](#)
- Luck, S. J. (2012). Event-related potentials. [38](#)
- Luck, S. J., Woodman, G. F., and Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in cognitive sciences*, 4(11):432–440. [39](#)
- Lundqvist, D., Flykt, A., and Öhman, A. (1998). The karolinska directed emotional faces (kdef). CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, (1998). [10](#)
- Maher, S., Ekstrom, T., Holt, D., Ongur, D., and Chen, Y. (2015). The core brain region for face processing in schizophrenia lacks face selectivity. *Schizophrenia bulletin*, 42(3):666–674. [177](#), [181](#)
- Maher, S., Mashhoon, Y., Ekstrom, T., Lukas, S., and Chen, Y. (2016). Deficient cortical face-sensitive n170 responses and basic visual processing in schizophrenia. *Schizophrenia research*, 170(1):87–94. [174](#), [175](#), [177](#)
- Mandal, M. K., Pandey, R., and Prasad, A. B. (1998). Facial expressions of emotions and schizophrenia: A review. *Schizophrenia bulletin*, 24(3):399. [181](#)
- Martinez-Conde, S., Macknik, S. L., and Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3):229–240. [1](#)
- Maurer, D., Le Grand, R., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences*, 6(6):255–260. [1](#), [24](#), [26](#)
- McKone, E. (2004). Isolating the special component of face recognition: peripheral identification and a mooney face. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1):181. [16](#), [18](#), [19](#)
- McKone, E., Crookes, K., Jeffery, L., and Dilks, D. D. (2012). A critical review of the development of face recognition: Experience is less important than previously believed. *Cognitive neuropsychology*, 29(1-2):174–212. [16](#)

- McKone, E. and Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? toward a new theory of holistic processing. *Psychonomic Bulletin & Review*, 16(5):778–797. [17](#), [22](#)
- Meaux, E. and Vuilleumier, P. (2016). Facing mixed emotions: Analytic and holistic perception of facial emotion expressions engages separate brain networks. *NeuroImage*, 141:154–173. [1](#), [3](#), [20](#), [21](#), [23](#), [40](#), [64](#)
- Mehoudar, E., Arizpe, J., Baker, C. I., and Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of vision*, 14(7):6–6. [12](#)
- Meinhardt-Injac, B., Persike, M., and Meinhardt, G. (2013). Holistic face processing is induced by shape and texture. *Perception*, 42(7):716–732. [2](#), [24](#), [165](#)
- Mielliet, S., Caldara, R., and Schyns, P. G. (2011). Local jekyll and global hyde: the dual identity of face identification. *Psychological science*, 22(12):1518–1526. [32](#)
- Mielliet, S., Vizioli, L., He, L., Zhou, X., and Caldara, R. (2013). Mapping face recognition information use across cultures. *Frontiers in psychology*, 4. [14](#), [32](#)
- Morris, R. W., Weickert, C. S., and Loughland, C. M. (2009). Emotional face processing in schizophrenia. *Current opinion in psychiatry*, 22(2):140–146. [12](#), [46](#), [175](#), [181](#)
- Murphy, D. and Cutting, J. (1990). Prosodic comprehension and expression in schizophrenia. *Journal of Neurology, Neurosurgery & Psychiatry*, 53(9):727–730. [188](#)
- Namdar, G., Avidan, G., and Ganel, T. (2015). Effects of configural processing on the perceptual spatial resolution for face features. *Cortex*, 72:115–123. [24](#)
- Neath, K. N. and Itier, R. J. (2014). Facial expression discrimination varies with presentation time but not with fixation on features: a backward masking study using eye-tracking. *Cognition & emotion*, 28(1):115–131. [29](#)
- Neath-Tavares, K. N. and Itier, R. J. (2016). Neural processing of fearful and happy facial expressions during emotion-relevant and emotion-irrelevant tasks: A fixation-to-feature approach. *Biological psychology*, 119:122–140. [12](#), [33](#), [173](#), [178](#)
- Negrini, M., Brkić, D., Pizzamiglio, S., Premoli, I., and Rivolta, D. (2017). Neurophysiological correlates of featural and spacing processing for face and non-face stimuli. *Frontiers in Psychology*, 8. [39](#)
- Nemrodov, D., Anderson, T., Preston, F. F., and Itier, R. J. (2014). Early sensitivity for eyes within faces: A new neuronal account of holistic and featural processing. *NeuroImage*, 97:81–94. [4](#), [41](#), [44](#), [173](#)

- Neta, M., Tong, T. T., Rosen, M. L., Enersen, A., Kim, M. J., and Dodd, M. D. (2017). All in the first glance: first fixation predicts individual differences in valence bias. *Cognition and Emotion*, 31(4):772–780. [30](#)
- Nyström, M. and Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods*, 42(1):188–204. [32](#)
- Öhman, A., Lundqvist, D., and Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. *Journal of personality and social psychology*, 80(3):381. [27](#)
- Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., and Ohme, R. K. (2014). Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. *Frontiers in psychology*, 5. [10](#)
- Palermo, R. and Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods*, 36(4):634–638. [3](#), [13](#), [20](#), [30](#), [164](#)
- Palermo, R. and Rhodes, G. (2007). Are you always on my mind? a review of how face perception and attention interact. *Neuropsychologia*, 45(1):75–92. [12](#), [14](#)
- Peters, J. C. and Kemner, C. (2017). Facial expressions perceived by the adolescent brain: towards the proficient use of low spatial frequency information. *Biological Psychology*. [45](#)
- Peterson, M. F. and Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, 109(48):E3314–E3323. [34](#), [35](#)
- Phillips, M. L., Bullmore, E. T., Howard, R., Woodruff, P. W., Wright, I. C., Williams, S. C., Simmons, A., Andrew, C., Brammer, M., and David, A. S. (1998). Investigation of facial recognition memory and happy and sad facial expression perception: an fmri study. *Psychiatry Research: Neuroimaging*, 83(3):127–138. [20](#)
- Piepers, D. W. and Robbins, R. A. (2012). A review and clarification of the terms “holistic,” “configural,” and “relational” in the face perception literature. *Frontiers in psychology*, 3. [1](#), [17](#), [18](#), [19](#)
- Pochedly, J. T., Widen, S. C., and Russell, J. A. (2012). What emotion does the “facial expression of disgust” express? *Emotion*, 12(6):1315. [13](#), [30](#)
- Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., and Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces:

- Time course and topographic evoked-potentials mapping. *Human brain mapping*, 26(1):65–79. [40](#), [41](#)
- Prkachin, G. C. (2003). The effects of orientation on detection and identification of facial expressions of emotion. *British journal of psychology*, 94(1):45–62. [25](#), [63](#)
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology*, 62(8):1457–1506. [29](#)
- Regan, D. (1989). Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine. [39](#)
- Rhodes, G. (2013). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception*, 42(11):1179–1199. [18](#)
- Richler, J. J., Cheung, O. S., and Gauthier, I. (2011a). Holistic processing predicts face recognition. *Psychological science*, 22(4):464–471. [33](#)
- Richler, J. J., Gauthier, I., Wenger, M. J., and Palmeri, T. J. (2008). Holistic processing of faces: perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2):328. [17](#)
- Richler, J. J., Mack, M. L., Palmeri, T. J., and Gauthier, I. (2011b). Inverted faces are (eventually) processed holistically. *Vision research*, 51(3):333–342. [7](#), [24](#)
- Richler, J. J., Wong, Y. K., and Gauthier, I. (2011c). Perceptual expertise as a shift from strategic interference to automatic holistic processing. *Current directions in psychological science*, 20(2):129–134. [24](#)
- Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta psychologica*, 128(2):274–289. [17](#), [22](#)
- Rossion, B. (2009). Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis. *Acta psychologica*, 132(3):300–312. [24](#)
- Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, 21(2):139–253. [41](#), [163](#)
- Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends in cognitive sciences*, 18(6):310–318. [3](#), [34](#), [35](#), [38](#), [113](#)
- Rossion, B. and Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of vision*, 8(4):3–3. [2](#), [170](#)
- Rossion, B., Curran, T., and Gauthier, I. (2002). A defense of the subordinate-level expertise account for the n170 component. *Cognition*, 85(2):189–196. [44](#)

- Rossion, B., Delvenne, J.-F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., and Guérit, J.-M. (1999). Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biological psychology*, 50(3):173–189. [3](#), [40](#), [41](#)
- Rossion, B., Dricot, L., Goebel, R., and Busigny, T. (2010). Holistic face categorization in higher order visual areas of the normal and prosopagnosic brain: toward a non-hierarchical view of face perception. *Frontiers in Human Neuroscience*, 4. [1](#), [16](#)
- Rossion, B. and Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and cognitive neuroscience reviews*, 1(1):63–75. [1](#), [24](#)
- Rossion, B. and Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? ten lessons on the n170. *Neuroimage*, 39(4):1959–1979. [42](#)
- Rossion, B., Joyce, C. A., Cottrell, G. W., and Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage*, 20(3):1609–1624. [40](#), [43](#)
- Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2008). Time course and robustness of erp object and face differences. *Journal of Vision*, 8(12):3–3. [170](#)
- Russell, R., Biederman, I., Nederhouser, M., and Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision research*, 47(2):157–165. [24](#), [25](#)
- Sadeh, B., Podlipsky, I., Zhdanov, A., and Yovel, G. (2010). Event-related potential and functional mri measures of face-selectivity are highly correlated: a simultaneous erp-fmri investigation. *Human brain mapping*, 31(10):1490–1501. [117](#)
- Sagiv, N. and Bentin, S. (2001). Structural encoding of human and schematic faces: holistic and part-based processes. *Journal of Cognitive Neuroscience*, 13(7):937–951. [1](#), [4](#), [15](#), [25](#), [39](#), [40](#), [41](#), [42](#), [43](#), [45](#), [171](#), [172](#)
- Scheller, E., Büchel, C., and Gamer, M. (2012). Diagnostic features of emotional expressions are processed preferentially. *PloS one*, 7(7):e41792. [1](#), [3](#), [12](#), [14](#), [32](#), [173](#), [178](#)
- Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., and Pfurtscheller, G. (2007). A fully automated correction method of eeg artifacts in eeg recordings. *Clinical neurophysiology*, 118(1):98–104. [178](#)
- Schurigin, M., Nelson, J., Iida, S., Ohira, H., Chiao, J., and Franconeri, S. (2014). Eye movements during emotion recognition in faces. *Journal of vision*, 14(13):14–14. [14](#), [15](#), [19](#), [23](#), [29](#), [30](#), [32](#), [33](#), [34](#), [36](#), [66](#), [164](#), [166](#), [173](#)

- Schwaninger, A., Wallraven, C., Cunningham, D. W., and Chiller-Glaus, S. D. (2006). Processing of facial identity and expression: a psychophysical, physiological, and computational perspective. *Progress in Brain Research*, 156:321–343. [17](#)
- Schwartz, B. L., Marvel, C. L., Drapalski, A., Rosse, R. B., and Deutsch, S. I. (2002). Configural processing in face recognition in schizophrenia. *Cognitive neuropsychiatry*, 7(1):15–39. [176](#), [181](#)
- Schwarzer, G., Huber, S., and Dümmler, T. (2005). Gaze behavior in analytical and holistic face processing. *Memory & Cognition*, 33(2):344–354. [34](#)
- Schyns, P. G., Bonnar, L., and Gosselin, F. (2002). Show me the features! understanding recognition from the use of visual information. *Psychological science*, 13(5):402–409. [12](#), [177](#)
- Schyns, P. G. and Oliva, A. (1999). Dr. angry and mr. smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69(3):243–265. [2](#)
- Schyns, P. G., Petro, L. S., and Smith, M. L. (2009). Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: behavioral and brain evidence. *Plos one*, 4(5):e5625. [12](#), [14](#), [19](#), [34](#), [41](#)
- Sekiguchi, T. (2011). Individual differences in face memory and eye fixation patterns during face learning. *Acta psychologica*, 137(1):1–9. [12](#)
- Shin, Y.-W., Na, M. H., Ha, T. H., Kang, D.-H., Yoo, S.-Y., and Kwon, J. S. (2007). Dysfunction in configural face processing in patients with schizophrenia. *Schizophrenia bulletin*, 34(3):538–543. [189](#)
- Shipley, T. and Kellman, P. (2001). Gaze control for face learning and recognition by humans and machines. *From fragments to objects: Segmentation and grouping in vision*, page 463. [29](#)
- Skiba, D. J. (2016). Face with tears of joy is word of the year: are emoji a sign of things to come in health care? *Nursing education perspectives*, 37(1):56–57. [7](#), [8](#)
- Smith, M. L., Cottrell, G. W., Gosselin, F., and Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological science*, 16(3):184–189. [3](#), [7](#), [12](#), [34](#)
- Spezio, M. L., Adolphs, R., Hurley, R. S., and Piven, J. (2007). Analysis of face gaze in autism using “bubbles”. *Neuropsychologia*, 45(1):144–151. [2](#), [21](#), [30](#)
- Tanaka, J. W. and Curran, T. (2001). A neural basis for expert object recognition. *Psychological science*, 12(1):43–47. [44](#)

- Tanaka, J. W. and Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology*, 46(2):225–245. 2, 17, 33
- Tanaka, J. W. and Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & cognition*, 25(5):583–592. 33
- Tanaka, J. W. and Simonyi, D. (2016). The “parts and wholes” of face recognition: A review of the literature. *The Quarterly Journal of Experimental Psychology*, 69(10):1876–1889. 1, 18
- Taylor, M., Batty, M., and Itier, R. (2004). The faces of development: a review of early face processing over childhood. *Journal of cognitive neuroscience*, 16(8):1426–1442. 40
- Thompson, P. (1980). Margaret thatcher: a new illusion. *Perception*. 25
- Tomkins, S. (1962). *Affect imagery consciousness: Volume I: The positive affects*. Springer publishing company. 8
- Tomkins, S. S. and McCarter, R. (1964). What and where are the primary affects? some evidence for a theory. *Perceptual and motor skills*, 18(1):119–158. 8
- Tzschaschel, E. A., Persike, M., and Meinhardt-Injac, B. (2014). The effect of texture on face identification and configural information processing. *Psihologija*, 47(4):433–447. 2, 23
- Vaidya, A. R., Jin, C., and Fellows, L. K. (2014). Eye spy: The predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition*, 133(2):443–456. 14, 19, 30, 31, 32, 33, 34, 66, 168
- Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology*, 79(4):471–491. 24
- Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., and Lefèvre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision*, 10(5):10–10. 16, 18
- Vuilleumier, P. and Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia*, 45(1):174–194. 3, 20, 39, 40
- Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of nonverbal behavior*, 17(1):3–28. 58
- Walker-Smith, G. J. (1978). The effects of delay and exposure duration in a face recognition task. *Attention, Perception, & Psychophysics*, 24(1):63–70. 12

- Walker-Smith, G. J., Gale, A. G., and Findlay, J. M. (2013). Eye movement strategies involved in face perception. *Perception*, 42(11):1120–1133. [11](#), [12](#), [29](#)
- Wallace, S., Coleman, M., and Bailey, A. (2008). An investigation of basic facial expression recognition in autism spectrum disorders. *Cognition and Emotion*, 22(7):1353–1380. [12](#), [21](#), [28](#)
- Waller, B. M. and Micheletta, J. (2013). Facial expression in nonhuman animals. *Emotion Review*, 5(1):54–59. [8](#)
- Wang, H., Sun, P., Ip, C., Zhao, X., and Fu, S. (2015). Configural and featural face processing are differently modulated by attentional resources at early stages: An event-related potential study with rapid serial visual presentation. *Brain research*, 1602:75–84. [17](#)
- Watanabe, S., Kakigi, R., and Puce, A. (2003). The spatiotemporal dynamics of the face inversion effect: a magneto-and electro-encephalographic study. *Neuroscience*, 116(3):879–895. [176](#)
- Watson, T. L. (2013). Implications of holistic face processing in autism and schizophrenia. *Frontiers in psychology*, 4. [175](#), [181](#)
- Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., and Kissler, J. (2017). Mapping the emotional face. how individual face parts contribute to successful emotion recognition. *PloS one*, 12(5):e0177239. [11](#), [12](#), [64](#), [163](#), [164](#), [165](#)
- Whalen, P. J. (2004). Human Amygdala Responsivity to Masked Fearful Eye Whites. *Science*, 306(5704):2061–2061. [27](#), [63](#), [163](#), [173](#)
- Wiese, H., Schweinberger, S. R., and Hansen, K. (2008). The age of the beholder: Erp evidence of an own-age bias in face memory. *Neuropsychologia*, 46(12):2973–2985. [112](#)
- Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, 15(8):1234–1242. [44](#)
- Xu, Y., Liu, J., and Kanwisher, N. (2005). The m170 is selective for faces, not for expertise. *Neuropsychologia*, 43(4):588–597. [44](#)
- Yang, E., Zald, D. H., and Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion*, 7(4):882. [176](#)
- Yarbus, A. L. (1967). *Eye movements during perception of complex objects*. Springer. [11](#), [29](#)
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, 81(1):141. [2](#), [24](#)

- Yin, R. K. (1970). Face recognition by brain-injured patients: a dissociable ability? *Neuropsychologia*, 8(4):395–402. [24](#)
- Zeman, P. M., Till, B. C., Livingston, N. J., Tanaka, J. W., and Driessen, P. F. (2007). Independent component analysis and clustering improve signal-to-noise ratio for statistical analysis of event-related potentials. *Clinical Neurophysiology*, 118(12):2591–2604. [113](#)
- Zhao, M., Bülthoff, H. H., and Bülthoff, I. (2016). A shape-based account for holistic face processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(4):584. [2](#), [22](#), [23](#), [24](#), [165](#)

# Chapter 6

## Résumé / Abstract

**Titre :** Contribution des caractéristiques diagnostiques dans la reconnaissance des expressions faciales émotionnelles : une approche neurocognitive alliant oculométrie et électroencéphalographie

**Mots clés :** Emotion, Traitement du visage, Potentiels évoqués (ERPs), Eye-tracking, schizophrénie

**Résumé :** La reconnaissance experte de l'expression faciale est cruciale pour l'interaction et la communication sociale. Le comportement, les potentiels évoqués (ERP), et les techniques d'oculométrie peuvent être utilisés pour étudier les mécanismes cérébraux qui participent au traitement visuel automatique. La reconnaissance d'expressions faciales implique non seulement l'extraction d'informations à partir de caractéristiques faciales diagnostiques, stratégie qualifiée de traitement local, mais aussi l'intégration d'informations globales impliquant des traitements configuratifs. Des nombreuses recherches concernant le traitement des informations faciales émotionnelles il apparaît que l'interaction des traitements locaux et configuratifs pour la reconnaissance des émotions est mal comprise. La complexité inhérente à l'intégration de l'information faciale est mise en lumière lorsque l'on compare la performance de sujets sains et d'individus atteints de schizophrénie, car ces derniers ont tendance à s'attarder sur quelques éléments locaux, parfois peu informatifs. Les différentes façons d'examiner les visages peuvent avoir un impact sur la capacité socio-cognitive de reconnaître les émotions. Pour ces raisons, cette thèse étudie le rôle des caractéristiques diagnostiques et configuratives dans la reconnaissance de l'expression faciale. En plus des aspects comportementaux, nous avons donc examiné la dynamique spatiale et temporelle des fixations à l'aide de mesures oculométriques, ainsi que l'activité électrophysiologique précoce considérant plus particulièrement les composantes P100 et N170. Nous avons créé de nouveaux stimuli des esquisses par une transformation numérique de portraits photos en esquisses, pour des visages exprimant colère, tristesse, peur, joie ou neutralité, issus de la base Radboud Faces Database, en supprimant les informations de texture du visage et ne conservant que les caractéristiques diagnostiques (yeux et sourcils, nez, bouche).

Ces esquisses altèrent le traitement configural en comparaison avec les visages photographiques, ce qui augmente le traitement des caractéristiques diagnostiques par traitement élémentaire, en contrepartie. La comparaison directe des mesures neurocognitives entre les esquisses et les visages photographiques exprimant des émotions de base n'a jamais été testée, à notre connaissance. Dans cette thèse, nous avons examiné (i) les fixations oculaires en fonction du type de stimulus, (ii) la réponse électrique aux manipulations expérimentales telles que l'inversion et la déconfiguration du visage.

Concernant, les résultats comportementaux montrent que les esquisses de visage transmettent suffisamment d'information expressive (compte tenu de la présence des caractéristiques diagnostiques) pour la reconnaissance des émotions en comparaison des visages photographiques. Notons que, comme attendu, il y avait un net avantage de la reconnaissance des émotions pour les expressions heureuses par rapport aux autres émotions. En revanche, reconnaître des visages tristes et en colère était plus difficile. Ayant analysé séparément les fixations successives, les résultats indiquent que les participants ont adopté un traitement plus local des visages croqués et photographiés lors de la deuxième fixation. Néanmoins, l'extraction de l'information des yeux est nécessaire lorsque l'expression transmet des informations émotionnelles plus complexes et lorsque les stimuli sont simplifiés comme dans les esquisses. Les résultats de l'électroencéphalographie suggèrent également que les esquisses ont engendré plus de traitement basé sur les parties. Les éléments transmis par les traits diagnostiques pourraient avoir fait l'objet d'un traitement précoce, probablement dû à des informations de bas niveau durant la fenêtre temporelle de la P100, suivi d'un décodage ultérieur de la structure faciale dans la fenêtre temporelle de la N170. En conclusion, cette thèse a permis de clarifier certains éléments de la

discussion concernant le traitement des visages par configuration et par partie pour la reconnaissance des émotions, et d'approfondir notre compréhension actuelle du rôle des caractéristiques diagnostiques et des informations configurales dans le traitement neurocognitif des expressions faciales des émotions.

**Title:** The contribution of diagnostic featural information to the recognition of emotion facial expressions: a neurocognitive approach with eye-tracking and electroencephalography

**Keywords:** Emotion, Face processing, Event-related potentials (ERPs), Eye-tracking, schizophrenia

**Abstract :** Proficient recognition of facial expression is crucial for social interaction. Behaviour, event-related potentials (ERPs), and eye-tracking techniques can be used to investigate the underlying brain mechanisms supporting this seemingly effortless processing of facial expression. Facial expression recognition involves not only the extraction of expressive information from diagnostic facial features, known as part-based processing, but also the integration of featural information, known as configural processing. Despite the critical role of diagnostic features in emotion recognition and extensive research in this area, it is still not known how the brain decodes configural information in terms of emotion recognition. Complexity of facial information integration becomes evident when comparing performance between healthy subjects and individuals with schizophrenia because those patients tend to process featural information on emotional faces. The different ways in examining faces possibly impact on social-cognitive ability in recognizing emotions. Therefore, this thesis investigates the role of diagnostic features and face configuration in the recognition of facial expression. In addition to behavior, we examined both the spatiotemporal dynamics of fixations using eye-tracking, and early neurocognitive sensitivity to face as indexed by the P100 and N170 ERP components.

In order to address the questions, we built a new set of sketch face stimuli by transforming photographed faces from the Radboud Faces Database through the removal of facial texture and retaining only the diagnostic features (e.g., eyes, nose, mouth) with neutral and four facial expressions - anger, sadness, fear, happiness. Sketch faces supposedly impair configural processing in comparison with photographed faces, resulting in increased sensitivity to diagnostic features through part-based processing. The direct comparison of

neurocognitive measures between sketch and photographed faces expressing basic emotions has never been tested. In this thesis, we examined (i) eye fixations as a function of stimulus type, and (ii) neuroelectric response to experimental manipulations such face inversion and deconfiguration. The use of these methods aimed to reveal which face processing drives emotion recognition and to establish neurocognitive markers of emotional sketch and photographed faces processing.

Overall, the behavioral results showed that sketch faces convey sufficient expressive information (content of diagnostic features) as in photographed faces for emotion recognition. There was a clear emotion recognition advantage for happy expressions as compared to other emotions. In contrast, recognizing sad and angry faces was more difficult. Concomitantly, results of eye-tracking showed that participants employed more part-based processing on sketch and photographed faces during second fixation. The extracting information from the eyes is needed when the expression conveys more complex emotional information and when stimuli are impoverished (e.g., sketch). Using electroencephalographic (EEG), the P100 and N170 components are used to study the effect of stimulus type (sketch, photographed), orientation (inverted, upright), and de-configuration, and possible interactions. Results also suggest that sketch faces evoked more part-based processing. The cues conveyed by diagnostic features might have been subjected to early processing, likely driven by low-level information during P100 time window, followed by a later decoding of facial structure and its emotional content in the N170 time window. In sum, this thesis helped elucidate elements of the debate about configural and part-based face processing for emotion recognition, and extend our current understanding of the role of diagnostic features and configural information during

neurocognitive processing of facial expressions  
of emotion.

