



Kent Academic Repository

**Pongakkasira, Kaewmart (2015) *Face Detection in Complex Natural Scenes*.
Doctor of Philosophy (PhD) thesis, University of Kent,.**

Downloaded from

<https://kar.kent.ac.uk/54792/> The University of Kent's Academic Repository KAR

The version of record is available from

This document version

UNSPECIFIED

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Face Detection in Complex Natural Scenes

A thesis submitted for the Degree of Ph.D. in the Faculty of Social Sciences at the
University of Kent

Kaewmart Pongakkasira

School of Psychology

University of Kent

September 2015

Abstract

Face detection is an important preliminary process for all other tasks with faces, such as expression analysis and person identification. It is also known to be rapid and automatic, which indicates that detection might utilise low-level visual information. It has been suggested that this consist of a 'skin-coloured, face-shaped template', while internal facial features, such as the eyes, nose and mouth might also help to optimise performance. To explore these ideas directly, this thesis first examined how shape and features are integrated into a detection template (Chapter 2). For this purpose, face content was isolated into three ranges of spatial frequency, comprising low (LSF), mid (MSF) and high (HSF) frequencies. Detection performance in these conditions was always compared with an original condition, which displayed unfiltered images in the full range of spatial frequency. Across five behavioural and eye-tracking experiments, detection was best for the original condition, followed by MSF, LSF and HSF faces. LSF faces, which provide only crude visual detail (i.e. gross colour shape), were detected as quickly as MSF faces but less accurate. In addition, LSF faces showed a clear advantage over HSF, which contains fine visual information (i.e. detailed lines of the eyes, nose, and mouth), in terms of detection speed and accuracy. These findings indicate that face detection is driven by simple information, such as the saliency of colour and shape, which supports the notion of a skin-coloured face-shape template. However, the fast and more accurate performance for faces in the full and mid-spatial frequencies also indicates that facial features contribute to optimize detection.

In Chapter 3, three further eye-tracking experiments are reported, which explore further whether the height-to-width ratio of a coloured-shape template might be important for detection. Performance was best when faces' natural height-to-width ratios were preserved compared to vertically and horizontally stretched faces. This indicates that this is an important element of the cognitive template for face template. The results also highlight that face detection differs from face recognition, which tolerates the same type of geometric disruption. Based on the results of Chapter 2 and 3, a model of face detection is proposed in Chapter 4. In this model, colour face-shape and features drive detection in parallel, but not necessarily at equal speed, in a "horse race". Accordingly, rapid detection is normally driven by salient colour and shape cues that preserve the height-to-width ratio of faces, but finer visual detail from features can facilitate this process when further information is needed.

Acknowledgements

I would like to thank Dr Markus Bindemann for his support and superb supervision throughout my PhD. I would also like to thank my families, particularly my mum and the *Kuhlmanns*, for their support and patience. This research was supported by the Royal Thai Government Scholarship program.

Declaration

I declare that this thesis is my own work carried out under the normal terms of supervision.

Kaewmart Pongakkasira

Publications

Chapter 2 (Experiments 1 to 5) of this thesis has been submitted for publication.

Chapter 3 (Experiments 6, 7 and 8) of this thesis has been published:

Pongakkasira, K., & Bindemann, M. (2015). The shape of the face template: Geometric distortions of faces and their detection in natural scenes. *Vision Research, 109*, 99-106.

Table of contents

ABSTRACT	2
ACKNOWLEDGEMENTS	4
CHAPTER 1 General Introduction	8
1.1 INTRODUCTION	9
1.2 VISUAL SEARCH	10
1.3 THE SPECIFICITY OF FACE DETECTION IN VISUAL SEARCH	11
1.4 FACTORS INFLUENCING FACE DETECTION	15
1.4.1 <i>Contextual information</i>	15
1.4.2 <i>Face colour and shape</i>	15
1.4.3 <i>Blurring and contrast reduction of faces</i>	16
1.4.4 <i>Luminance</i>	18
1.4.5 <i>Inversion and features</i>	19
1.5 A MODEL OF FACE DETECTION	21
1.6 SPATIAL FREQUENCY AND FACE PROCESSING	26
1.7 GEOMETRIC DISTORTIONS AND FACE PROCESSING	28
1.8 STRUCTURE OF THIS THESIS	29
CHAPTER 2 The Role of Spatial Frequency for Face Detection in Natural Scenes	32
INTRODUCTION	33
EXPERIMENT 1	35
EXPERIMENT 2	48
EXPERIMENT 3	55

EXPERIMENT 4	62
EXPERIMENT 5	67
CHAPTER 3 The Shape of The Face Template: Geometric Distortions of Faces and Their Detection in Natural Scenes	88
INTRODUCTION	89
EXPERIMENT 6	91
EXPERIMENT 7	100
EXPERIMENT 8	107
CHAPTER 4 Summary, Conclusions and Future Research	116
4.1 SUMMARY AND CONCLUSIONS	117
4.2 LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH	131
REFERENCES	134
APPENDIX	149

Chapter 1:

General Introduction

1.1 Introduction

Human face detection is the process by which observers find faces within the visual environment (see, e.g. Lewis & Edmonds, 2005; Lewis & Ellis, 2003; Tsao & Livingstone, 2008). This process appears to be distinct from subsequent categorization tasks (Bindemann & Lewis, 2013). However, in contrast to other tasks with faces, such as identification (see, e.g. Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990; Burton, Jenkins, Hancock, & White, 2005) and matching (e.g. Burton, White, & McNeill, 2010; Clutterbuck & Johnston, 2002; Johnston & Bindemann, 2013), emotion recognition (e.g. Calder, Burton, Miller, Young, & Akamatsu, 2001; Calder & Young, 2005), or gaze perception (e.g. Bayliss, di Pellegrino, & Tipper, 2004; Driver et al., 1999; Jenkins, 2007), face detection has been studied comparatively little in Psychology. This is surprising considering that detection is an important first step for all other tasks with faces.

In this thesis, the detection of faces in natural scenes is explored across two themes. The first theme explores how spatial frequency affects detection, to determine the nature of the visual information in a face that is utilized for this purpose. The second theme then examines how face shape might contribute to detection, by manipulating the height-to-width ratio of faces. I begin by outlining the principles of visual search. This is followed by a review of the existing evidence on face detection. I end this chapter by describing the methodology of the current work.

1.2 Visual search

Face detection is essentially a *visual search* task, which requires observers to find a target in visual displays by matching an external stimulus to an internal template. According to feature integration theory, the search for a target requires the combination of separable features, such as colour, orientation and shape. During visual search some of these features may be shared with distractor items in a display (Treisman & Gelade, 1980; Treisman & Souther, 1985). If reaction times for a target do not increase with the number of distractors in a display, then search for the target characteristics is said to be parallel. This ‘pop-out’ effect can be found if a target looks distinct from the distractors, for example, when a red circle is embedded among yellow rectangles. The explanation for this effect is that the target can be located pre-attentively, resulting in very fast detection. If, on the other hand, reaction times taken increase linearly with the size of a search array’s size, then targets and distractors must share some important visual features and search is not parallel. Instead, focused attention is required to identify each of the displayed items and the search is said to be ‘serial’ (Treisman & Gelade, 1980; Wolfe, 1994).

A number of studies have applied such visual search paradigms to face detection to explore whether ‘pop-out’ exists (Hershler & Hochstein, 2005; Lewis & Edmonds, 2003, 2005). This research has shown that faces do not pop out when distractors share ‘face-like’ elements. For example, response time for detecting an upright face among inverted distractor faces has been shown to increase with set-size, indicating a serial search process (see Figure 1.1) (Brown, Huey & Findlay, 1997; Kuehn & Jolicoeus, 1994; Nothdurft, 1993). In turn, this effect is reduced when the distractors look less face-like. For example, an intact upright face target can be found more quickly among scrambled distractor faces (Kuehn & Jolicoeus, 1994), and

detection is faster still when faces are embedded among distinctive non-face objects (Hershler & Hochstein, 2005) or in scenes (Lewis & Edmonds, 2003, 2005).



Figure 1.1 An example from Nothdurft (1993). An upright face target is shown among nine inverted faces.

1.3 The specificity of face detection in visual search

Studies of visual search indicate that face detection is distinct from other types of stimuli in a number of ways. Firstly, faces can be found more rapidly than other types of stimuli, such as dog faces, cars and clocks (Hershler, Golan, Bentin, & Hochstein, 2010). The speed of search can be defined by dividing response times (milliseconds) by the number of to-be-searched items in displays, with speeds of 6 ms/item or less indicating “pop-out” (Treisman & Souther, 1985). The normal rate of face search in arrays of 9 to 16 elements appears to be constant at 3 ms/item (Lewis & Edmonds, 2005). Even in bigger search arrays of up to 64 items, faces can be detected at speeds of 6 ms/item (Hershler & Hochstein, 2005). By contrast, other types of stimuli such as animal faces, houses and cars exhibit search rates of between 17 and 28 ms/item when these are embedded among other non-face objects (Hershler & Hochstein, 2005) (see Figure 1.2). Searching for faces, then, is special in the sense that it does not appear to be affected to the same extent as non-face targets by the number of distractors in a display.

Further evidence for the speed of face detection comes from saccade choice tasks, in which observers have to determine whether a face is present in the left or right visual field. Under these conditions, saccades towards the side on which a face is present are initiated within 100 ms of stimulus onset. By contrast, saccades toward non-face targets such as animals or vehicles, require 120-130 ms (Crouzet, Kirchner, & Thorpe, 2010). During the free viewing of scenes that contain a person in either the left or right visual field, observers' first fixations also tend to be directed toward faces on 90% of trials (Fletcher-Watson, Findlay, Leekam, & Benson, 2008). Taken together, these results indicate a speed and spatial advantage for faces in detection.

A detection advantage for faces has also been observed in other contexts. For example, newborn babies already appear to shift attention preferentially to face-like targets within minutes of birth, even though their visual system is not fully developed (for reviews see, Macchi, Simion, & Umiltà, 2001; Simion, Farroni, Cassia, Turati, & Barba, 2002; Turati, Simion, Milani, & Umiltà, 2002). The face advantage also survives some neurological impairments. For example, patients with hemispatial visual neglect are more likely to detect faces in the neglected visual field than non-face targets (Vuilleumier, 2000). Moreover, some blindsight patients, who lack the ability to consciously detect visual stimuli in the affected hemifield, still report the presence of faces (Morris, de Gelder, Weiskrantz, & Dolan, 2001). This advantage in capturing attention and overcoming visual extinction suggests that face detection is efficient and automatic, possibly operating via innate representations (Nestor, Vettel, & Tarr, 2013).

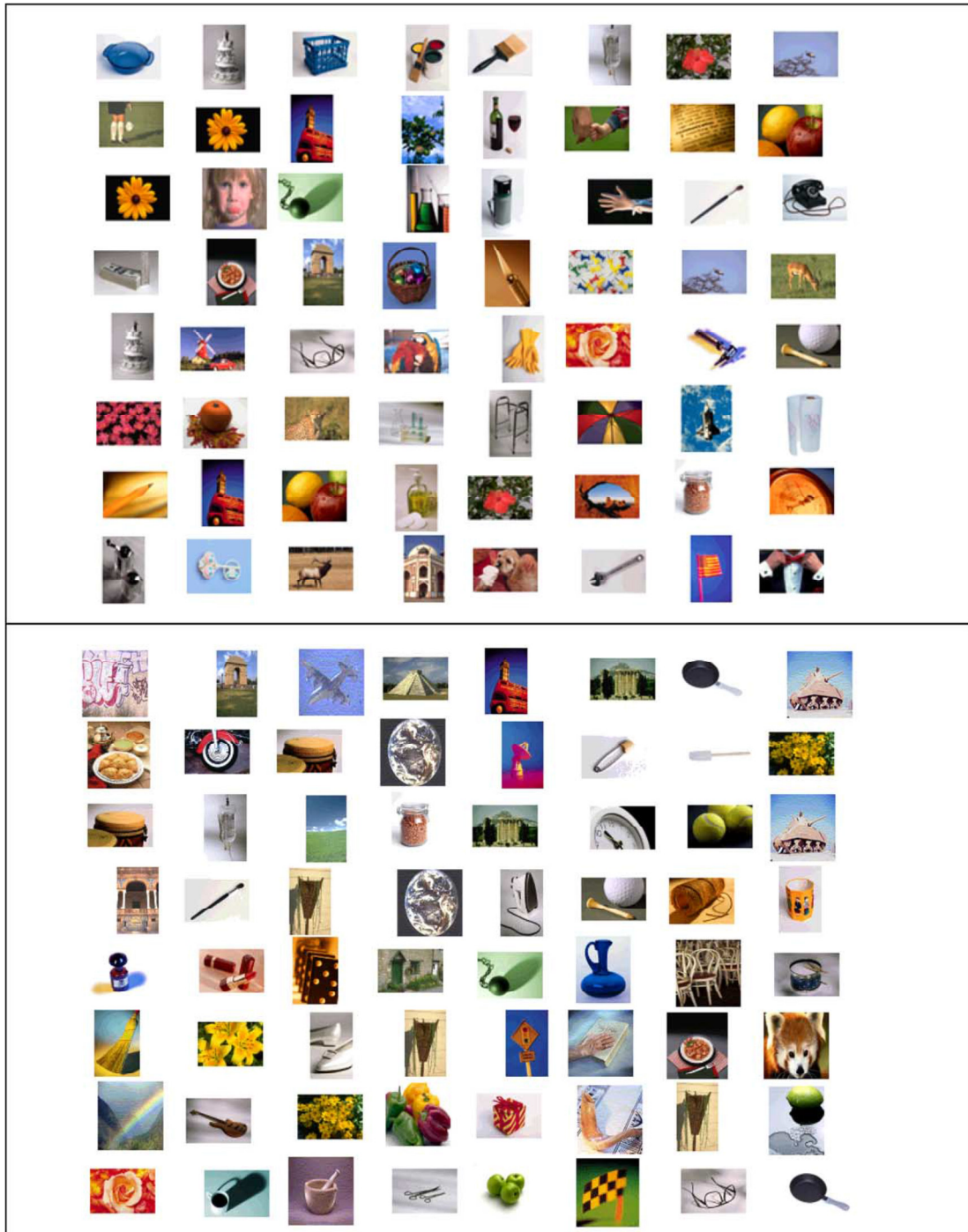


Figure 1.2 An example from Hershler and Hochstein (2005). A human face (top) and an animal face (bottom) target are shown among 64 items. Only mammal faces are used in this study to increase the similarity of features and configurations to human faces. Human faces appear to “pop out” from these displays, but animal faces do not.

Importantly, detection also appears to be distinct from other tasks with faces. For example, patients with prosopagnosia, which is an impairment in the ability to recognize familiar faces, can still detect faces as quickly as neurologically normal

subjects (Garrido, Duchaine, & Nakayama, 2008). When faces are embedded in complex visual scenes, a view effect is reliably found in neurologically normal observers, whereby profile faces are detected more slowly than frontal views (Burton & Bindemann, 2009). Crucially, however, this effect disappears when comparable face / non-face judgements are required to small, centrally-presented scenes, or when face and non-face objects are presented individually, without any extraneous background (Bindemann & Lewis, 2013) (see Figure 1.3). The difference between these tasks indicates that *search* for faces, in large displays, produces a response pattern that makes detection distinct from comparable face / non-face categorization tasks in which this search component is eliminated.

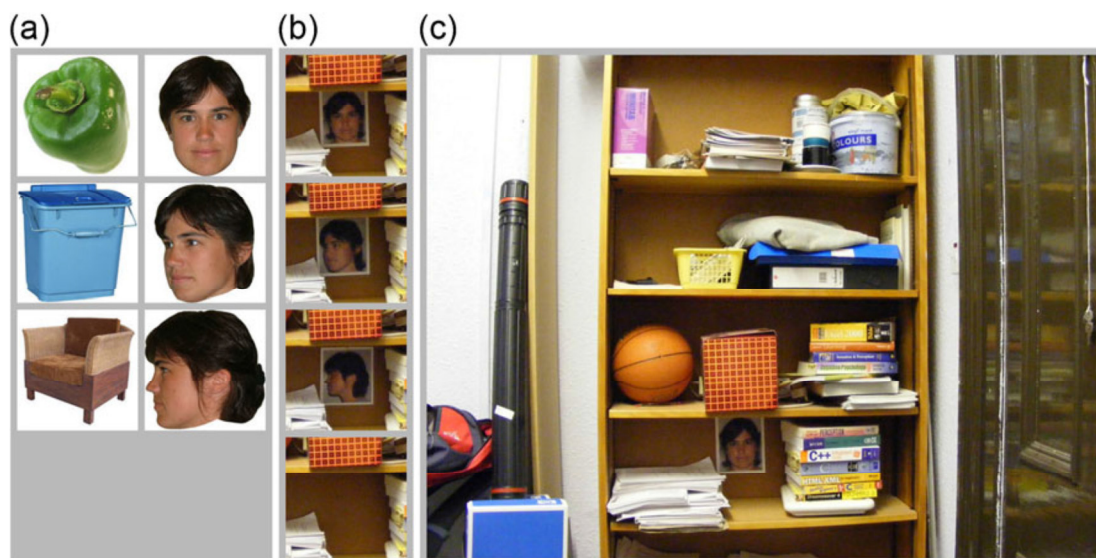


Figure 1.3 An example of isolated images of objects or faces (a), and the small (b) and large (c) scenes with faces used by Bindemann and Lewis (2013).

1.4 Factors influencing face detection

1.4.1 Contextual information

Face detection appears to be affected by a range of factors. When the scene context in which a face is presented is divided into rectangles and rearranged randomly (see Figure 1.4), detection performance declines (Lewis & Edmonds, 2003). This indicates that the surrounding information in a scene helps to guide observers towards faces or facilitates the decision as to whether an attended region contains a face or not. Blurring of visual scenes also impairs face detection, which provides further evidence that scenic information can direct attention to the region of a face. At the same time, blurring of to-be-search scenes facilitates 'absent' responses when faces are not present, which could also indicate that the reduction of scenic content can facilitate the scanning of visual displays for faces (Lewis & Edmonds, 2003).

1.4.2 Face colour and shape

Face detection appears to be facilitated by skin-colour and face-shape information (Bindemann & Burton, 2009; Lewis & Edmonds, 2003, 2005). Faces are detected faster in their veridical colours than when they are rendered in unnatural colours or greyscale (see Figure 1.5) (Bindemann & Burton, 2009). However, this advantage is only observed when skin-colour information is combined with face shape. When colour information is preserved in only part of a face, while the remaining area is rendered in greyscale, performance is comparable to faces that are presented entirely in greyscale (Bindemann & Burton, 2009). This indicates that skin colour and face shape operate in combination to facilitate detection.



Figure 1.4 An example from Lewis and Edmonds (2003), showing intact, scrambled and blurred scenes.

1.4.3 Blurring and contrast reduction of faces

Blurring and contrast reduction appear to reduce face detection in the manner that is comparable to the removal of colour. Blurring faces, for example with a Gaussian blur with a 3-pixel radius, or a contrast reduction of 50% increase detection times. Moreover, the effects of blurring and contrast reduction are super-additive,

which suggests that these factors disrupt the same processing stage (Lewis & Edmonds, 2003, 2005). It is possible that these factors actually do not affect detection during the initial search for faces, but at a subsequent decision-making stage, where observers have to decide whether a looked-at stimulus is in fact a face. This notion receives support from eye-tracking studies, which indicate that the reduction of image clarity increases detection decisions but not the delay the time to find a face target in first place (Awasthi, Friedman & Williams, 2011a, 2011b; Crouzet & Thorpes, 2011).

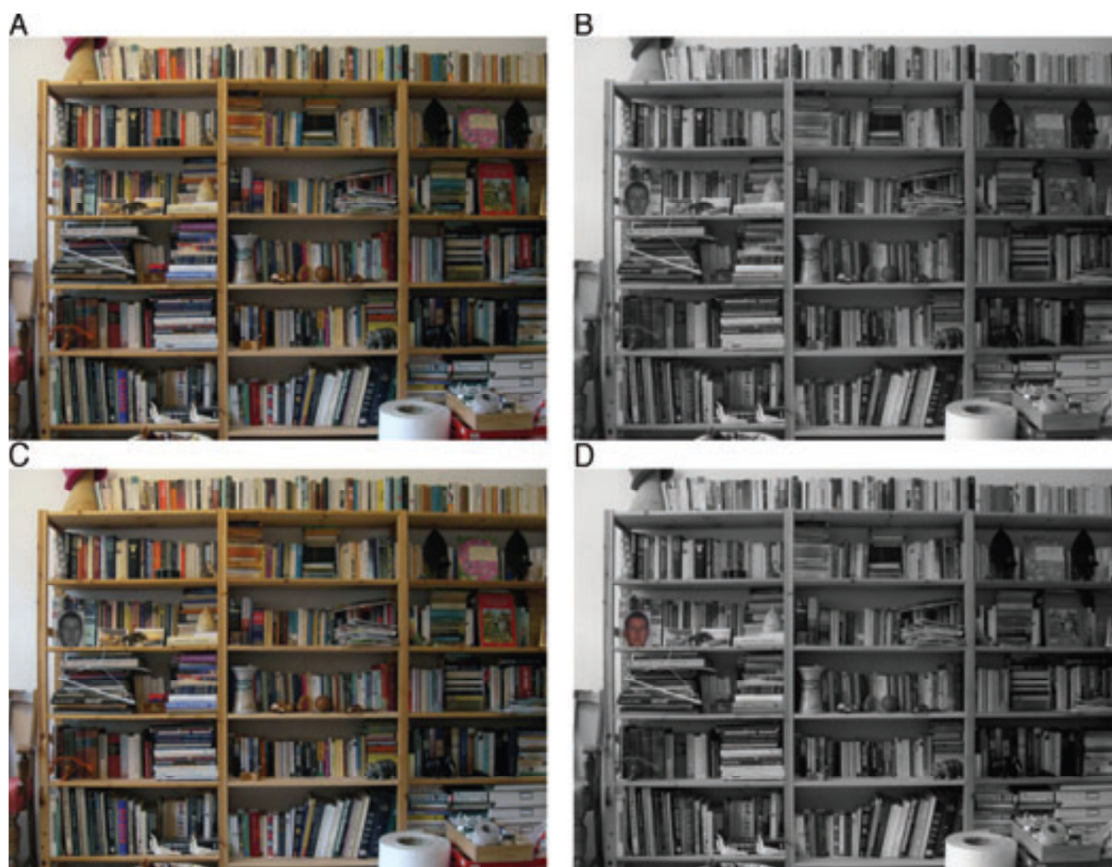


Figure 1.5 Example from Bindemann and Burton (2009) showing a face-absent scene in colour (A), a face-scene in greyscale (B), or with only the face (C) or the scene context (D) in greyscale.

In line with this reasoning, these factors have also been shown to reduce identification decisions that require similarity judgements between faces, such as matching tasks (Bindemann, Attard, Leach, & Johnston, 2013; Gilad, Meng, & Sinha, 2009). One way to reconcile these findings is that the blurred gross structure of faces

might be sufficient for guiding attention to possible face candidates in scenes. However, the reduction of image clarity by blurring and contrast reduction affects the finer evaluation of structure, such as edge extraction. If detection decisions rely on such detail for final template matching, then the reduction of such detail in blurred or contrast-reduced images might hamper detection.

1.4.4 Luminance

Luminance also appears to be important for face detection. The reversal of image luminance, whereby light areas are transposed into dark area by reversing pixel values, reduces the speed of face detection (Lewis & Edmonds, 2003, 2005), presumably by affecting shading from shape cues (Kemp, Pike, White, & Musselman, 1996). The shape of the head and facial features provide distinct patterns of light and dark that might form part of the template for face detection. For example, as the eyes are set in sockets, this concave inevitably provides a darker visual contrast of two horizontally-aligned circles in a face. If such light-dark contrasts are integrated into a detection template, then the disruption of this pattern through luminance reversal will reduce the match between a seen stimulus and observers' internal face template.

The eye regions might, in fact, be a particularly important feature in this context. When faces are reversed in brightness (by reversing the grey-level relationship in photographic negative images), recognition appears to be unaffected (Bruce & Langton, 1994). However, the specific brightness reversal of the eye regions impairs face encoding. In turn, these detrimental effects are eliminated when normal lightness is presented in the eye region alone (Kemp et al., 1996). The results suggest that shading from features is important for face encoding (Gilad, Meng, & Sinha, 2009). While luminance reversal affects face processing in this aspect, face detection

might not rely on the same mechanism as depth cue from shading of a particular feature might not improve detection performance (Bindemann & Lewis, 2013; Burton & Bindemann, 2009).

1.4.5 Inversion and features

Stimulus inversion, by turning images upside-down, also appears to impair face detection (see Figure 1.6) (Garrido, Duchaine, & Nakayama, 2008) and other localization tasks such as change detection (Ro, Russell, & Lavie, 2001). A possible explanation is that inversion disrupts template matching, by creating a mismatch between a stored internal detection template and an observed face stimulus. The mental rotation that is required to overcome this mismatch increases detection times. Another possibility is that there are separate templates for upright and inverted faces, but the latter is activated less frequently and therefore requires more time to activate.

There is considerable evidence for the idea of an upright template for face detection. Preferential tracking of simple face-like patterns, such as three dots that represent two eyes and a nose, is already evident in newborn infants (Johnson, Dziurawiec, Ellis, & Morton, 1991). Judgments about upright images that contain such a pattern also demonstrate an advantage in response speed in a face *localization* task, in which a target is briefly flashed either left or right of fixation (Purcell & Stewart, 1986, 1988). In turn, the disruption of this relationship through image scrambling or the interchanging of features, such as the nose and mouth, impairs detection (Garrido et al., 2008; Hershler & Hochstein, 2005).

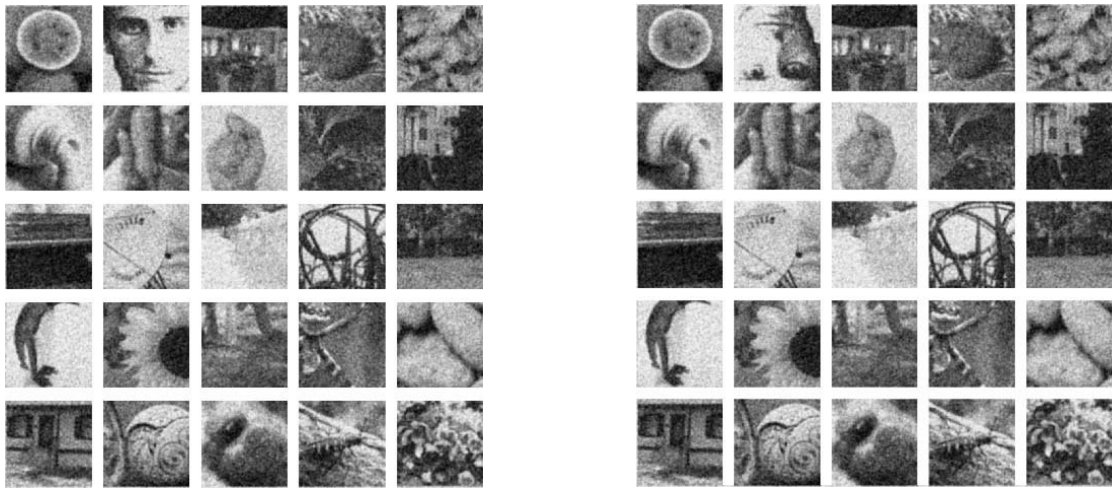


Figure 1.6 An example of visual search array from Garrido, Duchaine and Nakayama (2008). Faces are presented in upright (left) and inverted (right) orientation.

However, as both face shape and individual internal features are turned upside-down during inversion, it is unclear which of these aspects delays detection performance. The available evidence suggests that both types of information may give rise to this effect, depending on the circumstances under which faces are encountered. Some detection studies show no inversion effect when faces are presented in visual search arrays (Brown, Huey, & Findlay, 1997; Kuehn & Jolicoeur, 1994; Nothdurft, 1993) or only a small inversion effect of less than 20 ms (Lewis & Edmonds, 2003, 2005). Moreover, upright faces do not pop out when these are embedded among their inverted counterparts (Brown et al., 1997; Kuehn & Jolicoeur, 1994; Lewis & Edmonds, 2005; Nothdurft, 1993), and upright and inverted faces are equal competitors for observers' attention when they are presented together in the visual field (Bindemann & Burton, 2008) (Figure 1.7). This indicates that upright and inverted faces share visual characteristics that are important for detection, such as an oval skin-coloured shape.



Figure 1.7 Inversion has no effect in attention task (Bindemann & Burton, 2008). The stimuli are pairs of an upright face and an inverted face presented in the same trials. Subjects were told to fixate the centre of the display and then to make two-choice response according to the target's onscreen location.

On the other hand, there is also evidence that the internal facial features contribute to detection. Detection performance declines, for example, when face shape is preserved but the internal facial features are inverted (Macchi, Simion, & Umiltà, 2001; Olk & Garay-Vado, 2011) or scrambled (Valentine & Bruce, 1986). Similar to studies of luminance, a key feature in this context appears to be the eyes, as detection is impaired particularly when the eye regions are occluded (Lewis & Edmonds, 2003) or when only one, rather than both eyes, is visible (Burton & Bindemann, 2009). Moreover, detection is superior for the upper halves of faces, which contain the regions, than the lower half with the nose and mouth (Burton & Bindemann, 2009). However, while these findings point to the eyes as an important feature, face detection also appears to proceed unhindered in highly complex displays when all internal features (eyes, nose, mouth) are removed but a blank skin-coloured face shape is retained (Hershler & Hochstein, 2005).

1.5 A model of face detection

Based on the studies reviewed so far (see Table 1.1 for a summary), a possible model of face detection can be proposed to reconcile research on face-shape and facial features (see Figure 1.8). According to this model, general face-shape and salient global cues, such as colour, might help to identify possible face candidates within the

visual field. This is consistent with reports that skin-colour facilitates detection, but only when this is tied to face shape (Bindemann & Burton, 2009), and the finding that detection proceeds unhindered when all internal features are removed but a blank skin-coloured face shape is retained (Hershler & Hochstein, 2005). However, shape and colour alone cannot account for face detection, as faces in unnatural colours or greyscale are still detected well, albeit at a reduced speed (Bindemann & Burton, 2009). In addition, detection performance also declines when shape is disrupted through image scrambling (Hershler & Hochstein, 2005) or inversion (Garrido, Duchaine, & Nakayama, 2008). This suggests that, even though skin-colour and face-shape facilitate detection, additional cues support this process.

These additional cues are likely to be based on internal facial features. In the absence of skin-colour and global shape cues, internal features comprising the eyes, nose and mouth can still drive detection rapidly (Hershler & Hochstein, 2005). Indeed, even a simple configuration of four dots representing the eyes, nose and mouth appears sufficient to guide observers to face-like regions in the visual field (e.g. Johnson, Dziurawiec, Ellis, & Morton, 1991; Macchi, Simion, & Umiltà, 2001), and disruption of such information through inversion or scrambling delays detection (Garrido, Duchaine, & Nakayama, 2008). In addition, even simply ink blobs of black and white (Mooney faces), in which simple featural information is retained, can be detected as a face (Andrews & Schluppeck, 2004; George, Jemel, Fiori, Chaby, & Renault, 2005). This suggests that, in the absence of face-shape and colour, the internal facial features, arranged in a natural configuration (i.e. two eyes above a central nose and mouth), can also support detection.

It is hypothesized that our visual system utilizes these different types of visual information through ‘*template matching*’, by comparing a visual stimulus with a stored internal representation of a face. The overlap between stimulus and template, characterized by the shared visual information, then determines the speed and accuracy of detection (for similar ideas, see Valentine, 1991). For example, in this framework profile faces would be slower to detect (see Burton & Bindemann, 2009) because this view does not provide the full oval shape of frontal faces, due to the extended hair region across the head. In addition, fewer facial features are also observable in profile (see Bindemann, Scheepers, & Burton, 2009; Burton & Bindemann, 2009). This view might therefore provide a poor fit with a (frontal) template for face detection, both in terms of its shape and featural information.

At present, it is unresolved whether colour-shape information and internal features are processed in parallel or serially. In addition, it is possible that both types of information serve distinct purposes. For example, one intriguing possibility is that colour-shape information quickly helps to identify possible face candidates in the visual field. Once these face candidates are fixated, internal facial features, such as the eyes, might then be utilised in a decision process to confirm that a looked-at stimulus is, in fact, a face. This search-decision theory does not rule out that observers might also use features during search – for example, when colour-shape information is not readily available – but suggests that the primary function of such information might be confirmatory. In support of this idea, it is already known that the visual system is more effective at combining information from neurons that respond to the same visual characteristics, such as orientation and colour (Hubel & Wiesel, 1959; Sagi & Julesz, 1986), which facilitates processing. In addition, the visual system can alternate the integration of different types of information, such as blurred or detailed visual content

(see Johnson, 2005; Schyns & Oliva, 1997, 1999). For example, when seeing a person's face, the same neurons can convey two different types of facial information with different latencies, starting at coarse information (i.e. shape) followed by fine information (i.e. identity detail) (see, e.g. Halit, de Hann, Schyns, & Johnson, 2006; Sugase, Yumane, Ueno, & Kawano, 1999). If these principles apply also to face detection, then this could support a model in which colour-shape is primarily responsible for detection, but featural information is utilized for this purpose also when required.

In this thesis, I begin to explore these possibilities by measuring the speed and accuracy of face detection in natural scenes with observers' responses and eye movements. The reported experiments manipulate spatial frequency and, later on, the geometric dimensions of faces, to explore whether colour-shape or featural information is most likely to form the primary template for face detection.

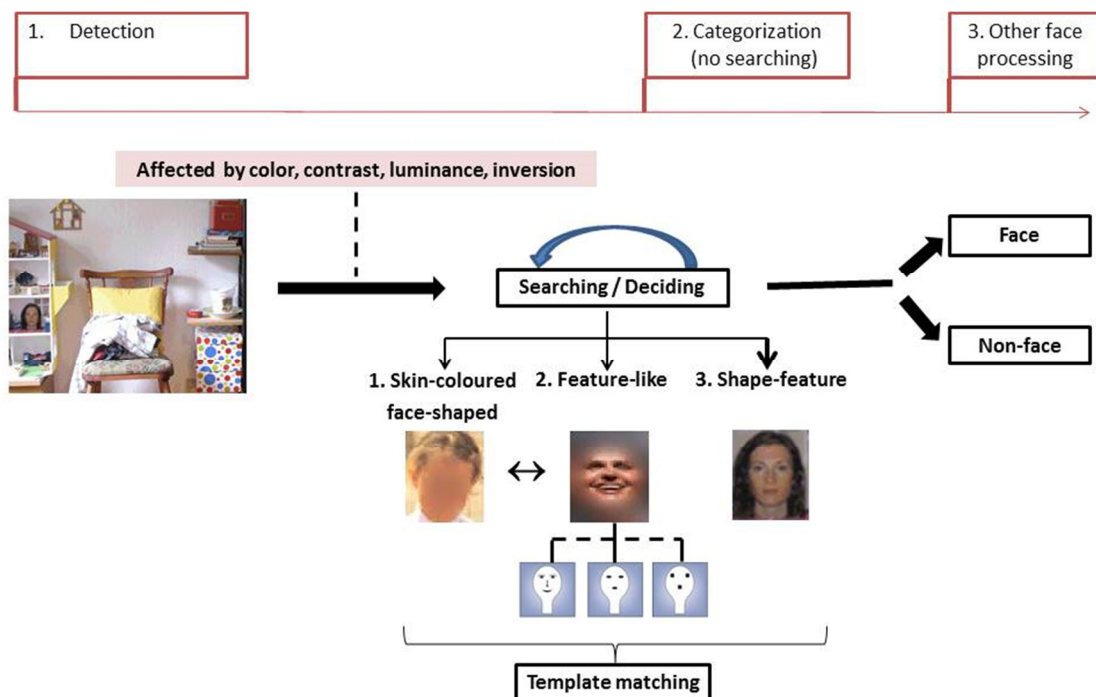


Figure 1.8 Proposed detection model. Possible face candidates are matched to internal templates are purposed, which might be based on skin-coloured face shapes, featural processing, and or shape-feature template that combines both.








References	Tasks	Results	Shape outline	Internal detail
1. Skin-coloured face-shaped template				
Bindemann & Burton (2009)	Detecting faces in scenes		✓	X
Bindemann & Lewis (2013)	Detecting faces in scenes		✓	X
Burton & Bindemann (2009)	Detecting faces in scenes		✓	X
2. Features-like template				
Johnson, Dziurawiec, Ellis, & Morton (1991)	Tracking of face-like		X	✓
Macchi, Simion, & Umiltà (2001)	Tracking of face-like		X	✓
Valentine & Bruce (1986)	Judgement of face structure		X	✓
3. Shape-Feature template				
Hershler & Hochstein (2005)	Detecting faces in visual arrays		✓	✓

Table 1.1 Summary of research evidence supporting face templates. The tasks shown employ detection or tracking paradigms, or involve judgement about face structure. In the summary of results ‘=’ stands for equal performance; ‘> <’ for better or lower performance; ‘✓’ underlines the importance of ‘shape outline’ or ‘internal detail’.

1.6 Spatial frequency and face processing

Any visual input can be broken down into patterns of light and dark called 'spatial frequency' (SF) (see Campbell & Robson, 1968; Westheimer, 2001). The role of spatial frequency relates to basic mechanisms of luminance extraction from visual input for tasks such as edge detection (Marr & Hildreth, 1980), movement (Morgan, 1992), and depth perception (Marshall, Burbeck, Ariely, Rolland, & Martin, 1996). Spatial frequency information can be broken down into several categories, each coding different aspects of visual stimuli. High spatial frequency (HSF) codes fine visual detail, such as lines, edges, and fine visual detail of stimuli, whereas gross shape information is carried by low spatial frequency (LSF). In addition, mid-spatial frequency (MSF) provides an intermediate level of detail between these two categories. The categories are defined by filtering information from images at different rates. For example, MSF information is typically sampled by applying Gaussian apertures that filter faces at bandwidths of between 8 and 16 cycles per face width (cycles/fw) (see, e.g. Costen, Parker, & Craw, 1996; Näsänen, 1999; Parker & Costen, 1999). LSF is extracted by applying filters of less than 8 cycles/fw, whereas HSF is extracted with more than 16 cycles/fw (see, e.g. Costen et al., 1996; Goffaux, Hault, Michel, Vuong, & Rossion, 2005).

At present, very little is known about the role of SF in face detection, but there is considerable research on the role of different SF in other face tasks. Face recognition, for example, appears to operate best on an intermediate level of detail that is coded by MSF, whereas the removal of this SF range impairs accuracy (see, e.g. Bachmann, 1991; Collin, Liu, Troje, McMullen, & Chaudhuri, 2004; Costen, Parker, & Craw, 1994, 1996; Morrison & Schyns, 2001; Näsänen, 1999; Ojanpää &

Näsänen, 2003; Tieger & Ganz, 1979). However, a number of studies also report that recognition is possible with pictures containing SF outside of MSF (Costen et al., 1996; Fiorentini, Maffei, & Sandini, 1983). For example, recognition memory for faces displayed in HSF has been found to be only 47 % accurate, compared to 90% for the original images (Davies, Ellis, & Shepherd, 1978). In contrast, recognition performance for blurred LSF faces appears to be unimpaired (Harmon & Julesz, 1973, Yip & Sinha, 2002).

Although recognition does not gain advantage from LSF and HSF, these bandwidths appear to be important for other categorization tasks. Sex decisions to faces, for example, rely predominantly on gross LSF information (Awasthi, Friedman, & Williams, 2011a; Schyns & Oliva, 1999). In contrast, emotion categorization appears to be driven by finer featural detail that is carried by HSF cues (Schyns, Bonnar, & Gosselin, 2002; Schyns & Oliva, 1999; Norman, & Ehrlich, 1987). In addition, judgements of faces' holistic properties (Goffaux & Rossion, 2006), configuration (the metric relations among features) (Goffaux et al., 2005), or orientation (Goffaux, Gauthier, & Rossion, 2003) appear to be driven by LSF. In contrast, featural processing, in tasks such as matching (Goffaux et al., 2005) or precise identification (Fiorentini, Maffei, & Sandini, 1983; Tieger & Ganz, 1979) seems to rely on HSF. Finally, LSF also appears to support the differentiation of faces and objects (Goffaux et al., 2003), especially in peripheral vision (Awasthi, Friedman, & Williams, 2011b).

With regard to face detection, these findings indicate that different SF bands might be useful for separable aspects of this task. Considering the advantage of LSF in lateral categorization (Awasthi et al., 2011b) and that this band codes gross image

information (see, e.g. Awasthi et al., 2011a, 2011b; Goffaux et al., 2005; Goffaux & Rossion, 2006; Schyns & Oliva, 1999), LSF might underlie the use of colour-shape information in detection. Thus, detection performance should be best when such information is available. By contrast, featural information is coded in the HSF band (Norman, & Ehrlich, 1987; Schyns et al., 2002; Schyns & Oliva, 1999). If this is not part of the primary template for face detection, then this should be impaired when only HSF information is preserved. Finally, MSF information might provide the best conditions for detection by providing an intermediate presentation that conveys some information about the gross structure of a face *and* also its features. The first aim of this thesis is to explore this directly.

1.7 Geometric distortions and face processing

If face detection is carried by gross visual information such as colour-shape, then the question arises as to the specific nature of this template. One characteristic that might be distinct in a colour-shape template is its dimensions such as the height-to-width ratio. While this has not been explored in face detection, research on face recognition has produced some surprising results. This work has shown that geometric distortions of faces, by stretching faces in a horizontal or vertical plane while the other dimension is retained, does not affect the accuracy or speed of recognition. This effect is remarkable in that it is found with dramatic transformations in which faces are stretched to 150% (Bindemann, Burton, Leuthold, & Schweinberger, 2008) or 200% (Hole, George, Eaves, & Rasek, 2002) of their original dimensions. Indeed, even neural responses to faces, such as the N250r, appear to be insensitive to these stretching manipulations (Bindemann et al., 2008). Moreover, this effect was found in a context in which the simple manipulation of stimulus inversion reduced recognition

accuracy and increased response times (Hole et al., 2002). Taken together, these findings indicate that height-to-width ratios are not important for face recognition.

However, it is unclear whether detection would be similarly tolerant to manipulations of height-to-width ratios. The purpose of face recognition is to distinguish different stimuli (i.e. individual identities) from the same category. This process relies on information that differentiates one person's face from another, and height-to-width ratio does not appear to be informative in this context. The purpose of face detection is different, as this process has to distinguish faces from non-face stimuli. Thus, whereas recognition has to operate upon information that is different across faces, detection has to operate on the similarities. At present, it is unresolved whether height-to-width ratio is sufficiently similar across faces to code such similarity. However, considering that face detection might be driven by a simple LSF colour-shape template, the question arises of what additional characteristics are preserved in such a stimulus. In this context, height-to-width ratio appears to be a plausible candidate. Thus, it will also be explored here.

1.8 Structure of this thesis

Recent studies suggest that a skin-coloured face-shaped template might be important for face detection (Bindermann & Burton, 2009; Hershler & Hochstein, 2005). However, evidence from face tracking (see Johnson, Dziurawiec, Ellis, & Morton, 1991; Macchi, Simion, & Umiltà, 2001) and categorization (Nestor, Vettel, & Tarr, 2013; Valentine & Bruce, 1986) also points to the involvement of internal facial features in this task. In addition, it remains unresolved whether this colour-shape and featural information is processed in parallel or might be used serially. For example,

colour-shape might be used to first find possible face candidates, whereas featural information might be used to confirm that a looked-at stimulus is, in fact, a face.

In this thesis, I will explore these ideas by asking observers to detect faces from different SF information. If colour-shape provides the primary cue to detect possible face candidates, then low-level information from LSF or MSF should facilitate detection the most. If, on the other hand, featural information is as important for this process, then faces should be detected as well from HSF, which selectively preserves such fine visual detail. The difference in detection speed and accuracy between such conditions should therefore provide insight into the importance of skin-coloured face-shaped and featural templates for face detection.

I begin to explore these ideas in Chapter 2 by presenting face photographs embedded in complex natural scenes. In the first three separate experiments faces are presented in an unfiltered (original) format or only LSF, MSF and HSF is preserved. Observers are then required to search these scenes to determine if a face is, in fact, present (i.e. make face-present versus face-absent decisions). To determine the usefulness of different SF, response times and accuracy are analysed. In addition, observers' eye movements are also tracked. The rationale for this additional measurement is that it might help to dissociate search processes for likely face candidates, as indexed by the eye movements that are required to first fixated a face in a scene, from subsequent decision processes to determine that a looked-at stimulus is a face, which should be reflected in observers' response times and accuracy. Thus, any differences between eye movement and response measure might help to dissociate these potentially serial processes.

These ideas are then explored further in additional two experiments whether the spatial frequency usage is actually linked to face processing or whether this might reflect other low-level processes in visual search. For this purpose, Experiment 4 explores how the removal of colour information affects SF usage in face detection. Experiment 5 then investigates whether SF provides salient cues in visual scenes that guide observers' attention irrespective of face content. This is investigated by selectively manipulating the SF information in a face or a correspondingly-size non-face region in scenes that are otherwise presented in an unfiltered format.

The final experimental chapter examines whether height-to-width ratios are an important component of the template for face detection in three further experiments. For this purpose faces are stretched in a vertical plane in Experiment 6 while the other dimension remains intact. The impact of this manipulation on detection is then explored by comparing it with unstretched faces. In Experiment 7, this manipulation is explored further by controlling the surface area of faces more precisely. Finally, Experiment 8 compares vertically and horizontally stretched faces.

Chapter 2:

The Role of Spatial Frequency for Face Detection in Natural Scenes

Introduction

Face detection is the process by which faces are noticed and located within the visual environment. This process appears to be distinct from other tasks with faces. For example, whereas detection is sensitive to changes in view (e.g. frontal versus profile faces), other categorization tasks, such as face / non-face decisions to stimuli at fixation, are not (Bindemann & Lewis, 2013). Face detection is also a very fast process that can be initiated within 100 ms of stimulus onset (Crouzet, Kirchner, & Thorpe, 2010). This speed suggests that face detection is driven by a “quick and dirty” processing strategy that is based on simple visual cues (Crouzet & Thorpe, 2011).

These superficial visual cues could reflect gross colour and shape information from faces. It has been shown, for example, that skin-colour tones facilitate detection, but only when they are tied to a full face-shape (Bindemann & Burton, 2009). Detection is also worse for profile views, in which the diagnostic oval shape of faces is disrupted naturally, than for frontal views (Burton & Bindemann, 2009). In contrast to these findings, the detail *within* a face appears to contribute little to detection. For example, detection appears to proceed unhindered when internal facial features, such as the eyes, nose and mouth, or external features, such as hairstyle, are removed as long as skin-colour and an oval face-shape is retained (Hershler & Hochstein, 2005).

Taken together, these studies suggest that the template for face detection might consist of a simple skin-coloured shape template that also preserves the general height-to-width ratio of faces. This indicates that detection is not driven by fine details but broader visual cues. In visual stimuli, these different cues are carried by a specific range on the luminance spectrum. On this spectrum, high spatial frequency (HSF) codes fine visual details, such as lines, edges, and fine visual detail of stimuli,

whereas gross shape information is carried by low spatial frequency (LSF). The available evidence from studies of face detection suggests that this process might be facilitated mostly by such LSF cues. To date, however, this idea has not been examined directly.

This is an interesting issue for two reasons. Firstly, the role of spatial frequency has already been explored in a wide range of face tasks. This research has revealed that different face tasks rely on distinct spatial frequencies. Sex decisions to faces, for example, appear to be based predominantly on gross LSF information, whereas emotion categorization is driven by the finer featural detail that is carried by HSF cues (Schyns, Bonnar, & Gosselin, 2002; Schyns & Oliva, 1999). Face recognition, on the other hand, appears to operate best on an intermediate level of detail that is coded by mid-spatial frequency (MSF) (see e.g. Bachmann, 1991; Collin, Liu, Troje, McMullen, & Chaudhuri, 2004; Costen, Parker, & Craw, 1994, 1996; Morrison & Schyns, 2001; Näsänen, 1999; Ojanpää & Näsänen, 2003; Tieger & Ganz, 1979). Despite these differences, the role of spatial frequency in face detection has not been assessed directly.

The second reason is that the spatial frequency information that drives face detection could provide some clues as to the underlying neurological pathways for this task. Two separable channels are known to be selectively tuned to specific spatial frequency bands (Rolls, Baylis, & Leonard, 1985). LSF, carrying large-scale luminance variations (Goffaux, Gauthier, & Rossion, 2003; Goffaux, Jemel, Jacques, Rossion, & Schyns, 2003), are carried by a fast, subcortical, magnocellular channel (Bullier, 2001; Livingstone & Hubel, 1988). In contrast, the small-scale luminance variations that are represented by HSF, and support the analysis of finer visual detail

(Schyns & Oliva, 1999), are processed by a comparatively slower parvocellular pathway (Bullier, 2001; Livingstone & Hubel, 1988). Thus, an investigation of the spatial frequency information that drives face detection might also reveal which of these pathways is most likely to subserve this process.

To explore these possibilities, the current study examined the role of spatial frequency in face detection over five experiments. In these experiments, faces were either presented with the full-range of spatial frequency intact or were filtered to selectively preserve low, mid or high spatial frequency content. The effect of these manipulations on the speed and accuracy with which faces can be detected in complex natural scenes was then examined.

Experiment 1

In Experiment 1, observers searched complex natural scenes for frontal views of faces. These scenes were either presented unfiltered, to display the full range of spatial frequency information, or were filtered to selectively preserve only the low, mid or high spatial frequency content. If face detection is driven by a gross face-shaped template, in which fine visual detail is not preserved, then performance should be best with LSF and the original scenes (which also contain LSF). If, on the other hand, visual detail also facilitates face detection, then the high detail of HSF (and the original scenes) might prove most useful for this purpose. Finally, it is also possible that the coarse detail of MSF, which provides an intermediate level of detail between the LSF and HSF ranges, is the best performance-match to the unfiltered original face stimuli.

Method

Participants

Twenty-four students (21 females) from the University of Kent, with a mean age of 23 years ($SD = 5.6$), participated in this experiment for course credit. All reported normal or corrected-to-normal vision.

Stimuli

The stimuli were adopted from previous detection studies (Bindemann, 2010; Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Burton & Bindemann, 2009) and consisted of 24-bit RGB photographs of 120 scenes, taken from inside houses, apartments and office buildings. These scenes measured 1000 (W) x 750 (H) pixels, and were presented at a resolution of 66 pixels/inch and a viewing distance of 60 cm. Two versions of each scene existed, which were identical in all aspects, except that one contained a photograph of a frontal face whereas the other did not.

The faces in the scene were unfamiliar faces of 20 different identities (10 males and 10 females). The size of these faces was varied across the scenes, ranging from 36 (H) x 27 (W) pixels to 139 (H) x 115 (W) pixels (mean dimensions and SD : $58.7 (\pm 19.4) \times 47.2$ pixels (± 16.2)). Thus, observers could not adopt a simple search strategy based on stimulus size. In addition, the scenes were also divided into an invisible 3 x 2 grid of six rectangular cells and the location of the faces was counterbalanced across these regions.

Three further versions were then produced of each of the face-present and face-absent scenes, which were Fourier-transformed to selectively preserve only the

LSF, MSF or HSF content. Based on previous discrimination tasks of face gender and expression (Aguado, Serrano-Pedraza, Rodríguez, & Román, 2010), cut-off values of less than 5 cycles/face and more than 15 cycles/face were chosen as low-pass and high-pass Gaussian filters to create the LSF and HSF conditions, while MSF were defined by the frequency bands between these two conditions. Applying this manipulation to all face-present and face-absent scenes (see Appendix A) resulted in a total of 960 displays, comprising 240 stimuli (120 face-present, 120 face-absent) for each of the original, LSF, MSF and HSF image conditions. Examples of face-present stimuli are shown in Figure 2.1 and 2.2.



Figure 2.1 An example of an original face-present scene in Experiment 1.

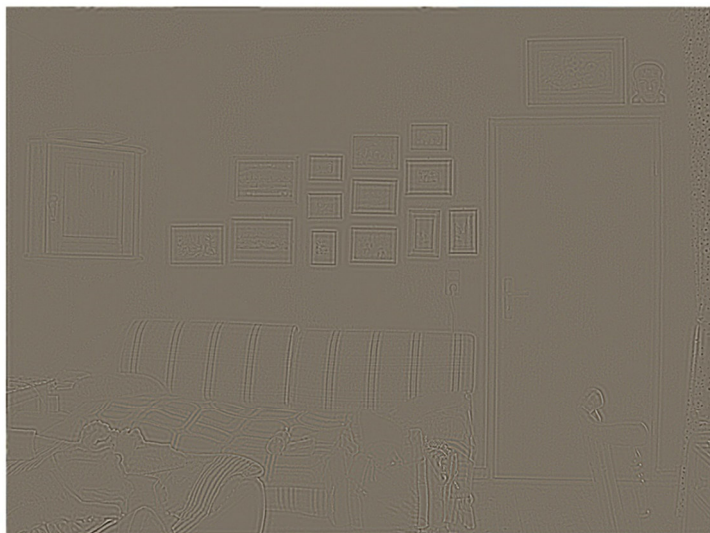


Figure 2.2 An example of a face-present scene in Experiment 1, depicting the LSF (top), MSF (middle), and HSF conditions (bottom).

Procedure

To measure visual search for faces directly, participants eye movements were tracked using an SR-Research Eyelink II head-mounted system running at 500 Hz sampling rate. Viewing was binocular but only the participants' dominant eye was tracked. To calibrate the eye-tracker, the standard 9-point Eyelink procedure was applied. Thus, participants fixated a series of nine targets on a 21 in. display monitor, which was positioned at a viewing distance of 60 cm. Calibration was then validated against a second presentation of these targets. If the latter indicated poor measurement accuracy (i.e. a mean deviation of more than 1° of participants estimated eye position from the target), calibration was repeated.

In the experiment, each trial began with an initial drift correction for which participants were required to focus on a central target. A scene stimulus was then shown until a response was registered. Participants were asked to decide whether a face was present or absent in the scene by pressing one of two possible buttons on a standard computer keyboard. Participants were informed in advance that the scenes would be manipulated to display different spatial frequency bands and might therefore appear blurry (e.g. in LSF) or consists of fine visual detail only (HSF).

A total of 360 trials was shown to each participant, consisting of 240 face-absent trials and 120 face-present trials, in a randomly intermixed order. For these conditions, 25% of the stimuli were shown in each of the original, LSF, MSF, and HSF format. The scene stimuli were rotated around these conditions, so that each face-present scene was only shown once, and each face-absent scene twice, to each participant in any of the conditions. Overall, however, the presentation of the scenes

was counterbalanced across participants, so that each scene appeared in each condition an equal number of times.

Results

Accuracy and response times

To assess detection performance, accuracy and the median correct reaction times were analysed for face-present scenes. The cross-subjects means of this data are illustrated in Figure 2.3 and show that accuracy was highest for the original scenes, followed by the MSF, HSF and LSF scenes. In line with these observations, a one-factor within-subject ANOVA of this data showed an effect of condition, $F(3,69) = 40.81$, $p < 0.001$, $\eta_p^2 = 0.64$. Post-hoc comparisons using Bonferroni t-tests were applied. To adjust for multiple comparisons, an alpha level of $p < 0.008$ was applied (i.e. for six comparisons, $p = 0.05/6$). Accuracy for the original condition was higher than for all other conditions (LSF, MSF, HSF), all $ts \geq 4.23$, $ps < 0.008$. Among the three spatial frequency conditions, accuracy was best for MSF, compared to LSF and HSF, both $ts \geq 5.72$, $ps < 0.008$, whereas the LSF and HSF conditions did not differ, $t(23) = 1.02$, $p = 0.32$.

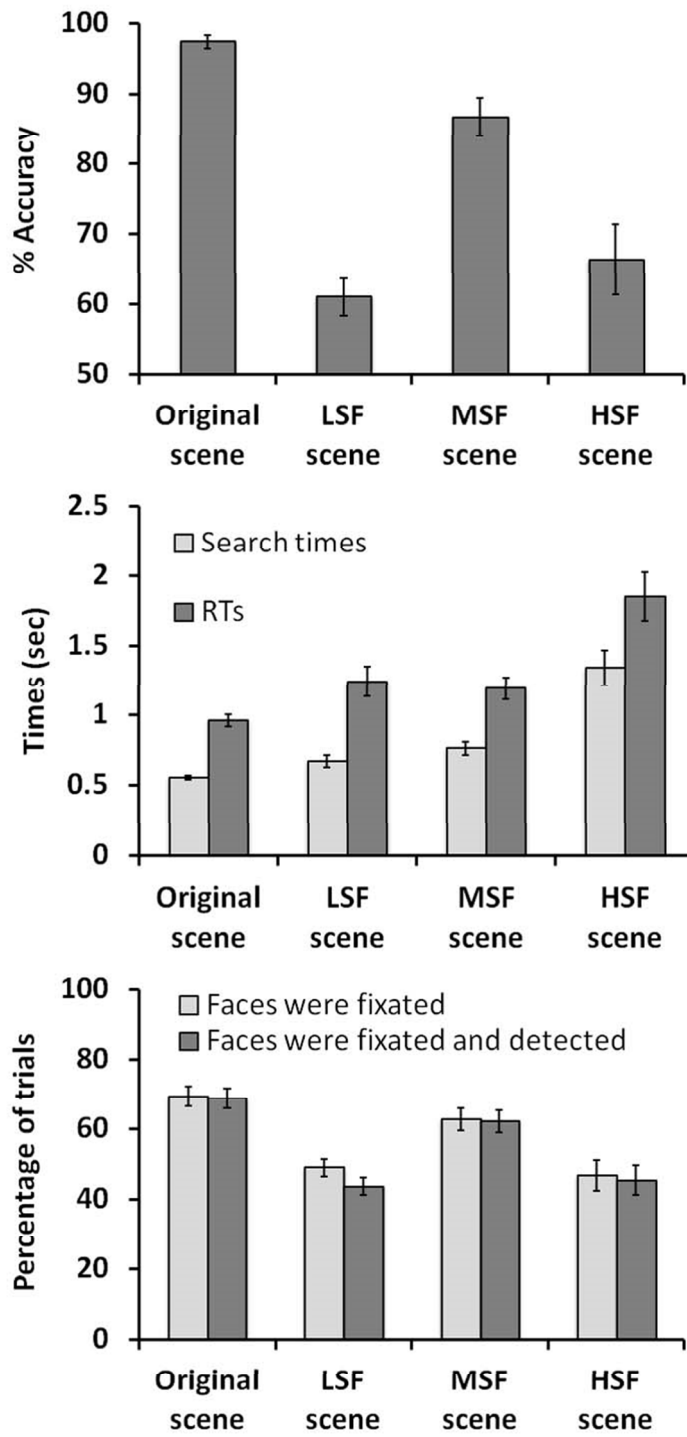


Figure 2.3 Detection performance for face-present scenes in Experiment 1, showing accuracy (top), reaction and search times (middle), and the eye movements to faces (bottom). Vertical bars represent the standard error of the means

Analysis of response times revealed a different pattern. As for accuracy, performance was best for the original scenes, but was similar for the MSF and LSF conditions, and lowest for HSF displays. These observations were confirmed by a one-factor within-subject ANOVA, $F(3,69) = 22.24$, $p < 0.001$, $\eta_p^2 = 0.49$. Bonferroni t-tests (with *alpha* corrected at $p < 0.008$) confirmed that faces were detected fastest in the original scene displays compared to all other conditions, all $ts \geq 3.12$, $ps \leq 0.008$, whereas performance was worst with HSF displays compared to all other conditions, all $ts \geq 3.95$, $ps \leq 0.008$. In contrast, performance was more evenly matched for MSF and LSF scenes, $t(23) = 0.49$, $p = 0.63$.

Eye movements

Eye movements were also processed to assess face detection across conditions. The percentage of trials on which faces were fixated in each condition was analysed first. Two measures are provided for this analysis. The first corresponds to mean of the percentage of trials on which faces were fixated *and* observers also made a correct face-present decision (i.e. all ‘fixated-and-detected’ trials). This measure essentially provides an eye-movement analogue to percentage accuracy (reported above). In addition, the mean percentage of trials on which faces were fixated is also reported. This includes all trials on which a face-absent response was erroneously made (i.e. all ‘fixated’ trials). This data is reported in Figure 2.3 and shows that these two measures were highly similar across all conditions. This indicates that faces that were fixated were typically also detected. In addition, these scores were highest for the original condition, followed by the MSF scenes, while performance was lowest and more comparable for the LSF and HSF conditions.

As these two measures are non-independent, they were analysed separately. For all fixated trials, a one-factor within-subject ANOVA showed a main effect of condition, $F(3,69) = 16.35$, $p < 0.001$, $\eta_p^2 = 0.42$. Post-hoc comparisons using Bonferroni t-tests (with *alpha* corrected at $p < 0.008$) showed that the percentage of these trials was comparable for the original and MSF faces, $t(23) = 2.01$, $p = 0.06$, which outperformed the LSF and HSF conditions, all $ts \geq 3.96$, $ps \leq 0.008$. In addition, performance for LSF and HSF faces also appeared to be closely matched, $t(23) = 0.48$, $p = 0.64$. The analogous analysis for the fixated-and-detected trials revealed a similar result. For this data, a one-factor within-subject ANOVA also showed an effect of condition, $F(3,69) = 20.02$, $p < 0.001$, $\eta_p^2 = 0.47$. The percentage of trials on which faces were fixated *and* detected was comparable for the original and MSF conditions, $t(23) = 2.07$, $p = 0.05$, and higher than for LSF and HSF displays, all $ts \geq 4.44$, $ps < 0.008$. The percentage of trials for LSF and HSF faces was again closely matched, $t(23) = 0.35$, $p = 0.73$.

In a next step, the eye movements were analysed to measure the time that was required to first fixate the faces in visual scenes on correct-response trials. This measure is included here to complement the response time data, but should provide a faster and more direct index of the search effort that is required to find a face. These *search times*, expressed as the mean of participants' medians, are also depicted in Figure 2.3 and correspond closely to the pattern of response times. A one-factor within-subject ANOVA showed a main effect of condition, $F(3,69) = 25.92$, $p < 0.001$, $\eta_p^2 = 0.53$, which reflects similar search times for faces in the original and LSF scenes, $t(23) = 2.81$, $p = 0.01$ (for *alpha* corrected at $p < 0.008$ for multiple comparisons), and faster search times for the original than MSF and HSF displays,

both $ts \geq 4.92$, $ps < 0.008$. In addition, performance was also comparable for LSF and MSF scenes, $t(23) = 1.96$, $p = 0.06$, and faster for both of these scene types than the HSF condition, both $ts \geq 4.74$, $ps < 0.008$.

Face-absent scenes

For completeness, observers' responses to face-absent scenes were also analysed (see Figure 2.4). A one-factor within subject ANOVA of the accuracy data showed an effect of condition, $F(3,69) = 8.17$, $p < 0.001$, $\eta_p^2 = 0.26$. Accuracy for original scenes was comparable to LSF and HSF scenes, both $ts \leq 2.45$, $ps \geq 0.02$ (with *alpha* corrected at $p < 0.008$ for multiple comparisons), and was also similar for LSF compared to MSF and HSF scenes, $t(23) \leq 2.42$, $ps \geq 0.02$. However, accuracy was higher for the original and HSF than MSF scenes, both $ts \geq 3.33$, $ps \leq 0.008$. A one-factor within-subject ANOVA of response times also showed a main effect of condition, $F(3,69) = 16.18$, $p < 0.001$, $\eta_p^2 = 0.41$, which reflects faster absent responses to LSF scenes than in the original, MSF and HSF conditions, all $ts \geq 4.61$, $ps < 0.008$. In contrast, performance for the original condition, MSF and HSF conditions was more similar, all $ts \leq 2.80$, $ps \geq 0.01$.

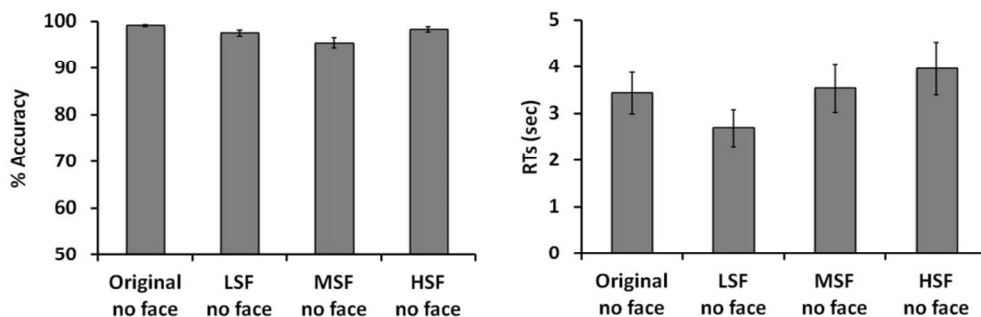


Figure 2.4 Detection performance for face-absent scenes in Experiment 1, showing accuracy (left panel) and reaction times (right panel). Vertical bars represent the standard error of the means.

Discussion

This experiment examined how spatial frequency supports face detection. For this purpose, detection was compared for faces embedded in original scenes, which retained all SF information, and scenes in which only LSF, MSF or HSF content was preserved. Detection accuracy was best for faces in the original and MSF scenes, and lowest for LSF and HSF scenes, which did not differ from each other. In contrast, faces were detected quickest in the original condition, but both MSF and LSF faces were detected faster than in the HSF condition. The findings are replicated in the eye movement data, which shows the same pattern in search times as the response times. The accuracy, response times and search times therefore converge by showing that detection is best for the original scenes, intermediate for MSF, and worst for HSF. However, these measures provide conflicting results for the LSF condition, in which faces are detected quickly but with low accuracy. Thus, this data indicates that LSF contains visual information that supports very fast detection – as fast as any SF bands and the unfiltered original faces. At the same time, it appears that this visual information also can be limiting in complex natural scenes and lead observers occasionally to miss faces entirely in the visual field.

A possible explanation for this discrepancy is that faces are actually located with greater accuracy from LSF than observers' responses suggest. This could occur if likely face candidates are located in this condition, but the LSF content provides insufficient detail to confirm that such a candidate region is actually a face. In this case, one would expect that faces are detected more often in the LSF condition than they are fixated *and* detected. In contrast to this notion, an analysis of observers'

fixations suggests that faces were fixated but *not* detected on only a small number of LSF trials (c.f. fixated and fixated-and-detected percentages in Figure 2.3).

An alternative explanation for the discrepancy in the accuracy and response times to faces in the LSF condition could reflect the fact that the entire scene stimuli were filtered to produce the different SF conditions in this experiment. Thus, the pattern of results might not reflect the effect that different SF have on *face* detection but might reflect scene processing instead. In line with this reasoning, it has already been shown that the perception of scene background is affected by SF, with blurred scenes producing faster absent responses (Lewis & Edmonds, 2003, 2005). In the current experiment, a similar pattern is evident in face-absent trials, in which responses were fastest in the LSF and slowest in the HSF conditions while both conditions were comparable for accuracy. Moreover, this pattern was similar to the face-present displays. These observations suggest that the effect of SF on detection might not reflect the processing of faces *per se*, but of the scene background. This was examined further in Experiment 2.

Before this is investigated, it is also notable that response times were generally slower on face-absent trials. This experiment adopted a contingency whereby face-absent scenes were presented twice as often as face-present scenes. Consequently, it is possible that the longer response times on face-absent trials reflect a belief in observers that more faces must be present (i.e., as would be the case with a 50:50 face present / absent ratio), which might increase the search effort, and therefore response times, on face-absent trials. While this is possible, reaction times are also longer in other studies that have used different target present to absent ratios (Lewis & Edmonds, 2003, 2005). An alternative explanation for the longer response on face-

absent trials is that search can be terminated for face-present conditions as soon as a face is located. Thus, this condition does not require that the entire scene is searched. If the face is not present, on the other hand, a more comprehensive search is necessary, causing a delay in response times.

Experiment 2

Experiment 1 demonstrates that face detection was fastest and most accurate for MSF, whereas HSF delayed and reduced the accuracy of detection. By contrast, LSF produced conflicting results by producing fast but inaccurate detection. One possible explanation for this finding is that these effects reflect the impact of SF on scene perception rather than face detection, as these manipulations were applied to the entire stimulus displays in Experiment 1. In turn, this raises the possibility that a different pattern is found when SF is manipulated in the face regions only, while the original, unfiltered scene background is retained. To examine this possibility, only the face photographs embedded within the scenes were filtered to display low, mid, and high spatial frequencies in Experiment 2.

Method

Participants

Twenty-one new undergraduate students (10 males, 11 females) from the University of Kent, with a mean age of 24 years ($SD= 3.2$), participated for course credit. All reported normal or corrected-to-normal vision.

Stimuli and procedure

The stimuli and procedure were identical to Experiment 1, except for the following changes. In this experiment, only one version of face-absent trials was retained from Experiment 1, which displayed the scenes in original SF content. For face-present trials, only the face photographs in the scenes were filtered to display low, mid, and high spatial frequencies, while the surrounding scene content was

unfiltered (i.e. original) (see Figure 2.5 and 2.6). This resulted in a total of 600 different displays comprising 120 face-absent scenes (all displaying original SF content) and 480 face-present images, in which faces were presented in original, LSF, MSF or HSF format. As in Experiment 1, each participant viewed 360 trials in a randomly intermixed order, comprising 240 face-absent trials and 120 face-present trials (30 images for each of the original, LSF, MSF, and HSF conditions). The face-present stimuli were rotated around these conditions so that each scene was only shown once to each participant in any of the conditions. Overall, however, the presentation of scenes was counterbalanced across participants, so that each scene appeared in an equal number of times in each condition.



Figure 2.5 An example of an original face-present scene in Experiment 2



Figure 2.6 An example of a face-present scene in Experiment 2, depicting the LSF (top), MSF (middle), and HSF conditions (bottom).

Results

Accuracy and response times

The mean accuracy and median correct reaction times are illustrated in Figure 2.7. For accuracy, a one-factor within-subject ANOVA showed an effect of condition, $F(3,60) = 37.70, p < 0.001, \eta_p^2 = 0.65$. Bonferroni t-tests (with *alpha* corrected at $p < 0.008$ for multiple comparisons) revealed that accuracy for original and MSF faces was similar, $t(20) = 2.07, p = 0.052$, and higher than for LSF and HSF faces, all $ts \geq 4.95, ps < 0.008$. In addition, accuracy for LSF faces was higher than for HSF faces, $t(20) = 4.12, p < 0.008$.

The pattern of response times complemented the accuracy data. Response times were fastest for the original condition, followed by MSF, LSF and HSF faces. A one-factor within-subject ANOVA of this data confirmed an effect of condition, $F(3,60) = 96.87, p < 0.001, \eta_p^2 = 0.83$. Bonferroni t-test (with *alpha* corrected at $p < 0.008$ for multiple comparisons) confirmed that response times were fastest for the original faces compared to all other conditions, all $ts \geq 5.75, ps < 0.008$. Responses to MSF and LSF did not differ significantly, $t(20) = 2.89, p = 0.009$, whereas HSF faces were detected slower than in any of the other conditions, all $ts \geq 9.40, ps < 0.008$.

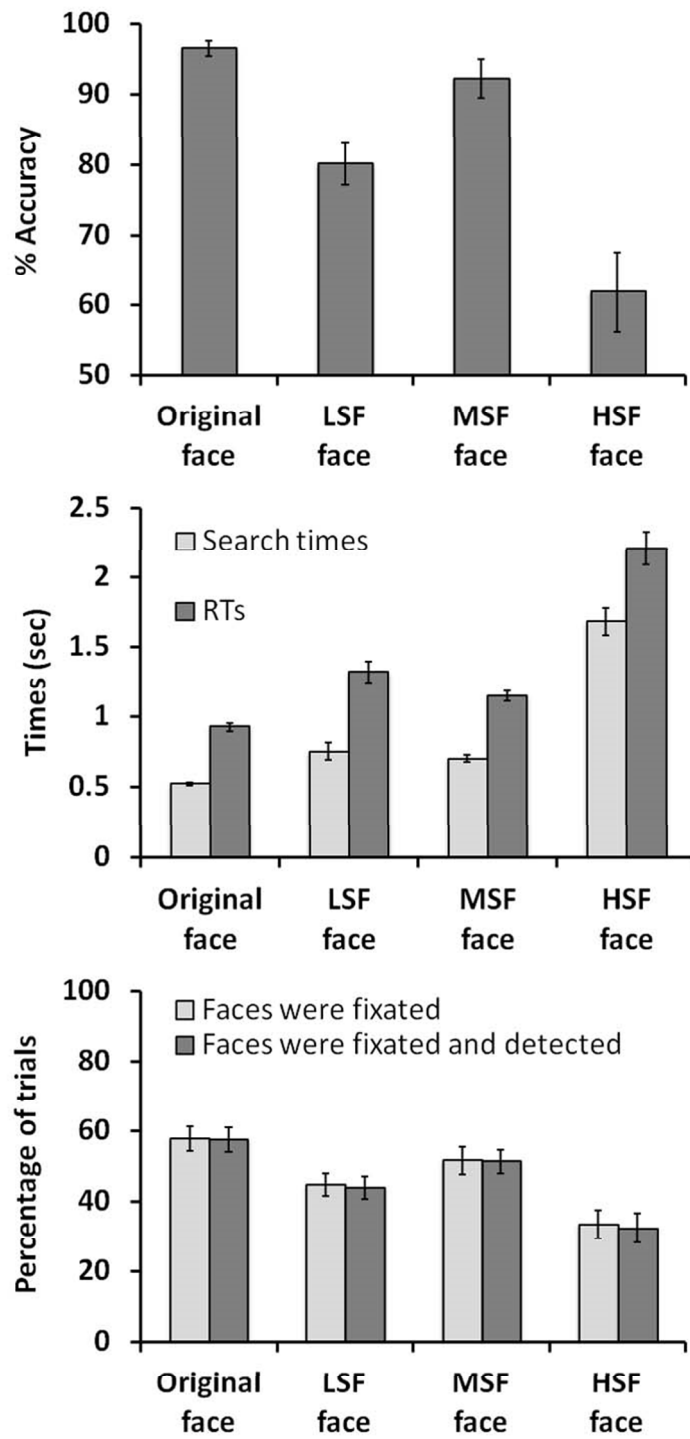


Figure 2.7 Detection performance for face-present scenes in Experiment 2, showing accuracy (top), reaction and search times (middle), and the eye movements to faces (bottom). Vertical bars represent the standard error of the means.

Eye movements

Eye movements were processed also to assess face detection across conditions. As in Experiment 1, the mean percentage of trials on which faces were fixated in each condition (all ‘fixated’ trials) and the percentage of trials on which faces were fixated *and* observers also made a correct face-present decision (i.e. all ‘fixated-and-detected’ trials) were analysed first. This data is reported in Figure 2.7 and shows that both measures were highly similar across all conditions. These scores were highest for the original condition, followed by the MSF, LSF and HSF faces.

For all fixated trials, a one-factor within-subject ANOVA showed a main effect of condition, $F(3,60) = 21.65$, $p < 0.001$, $\eta_p^2 = 0.52$. Bonferroni t-tests (with *alpha* corrected at $p < 0.008$ for multiple comparisons) revealed that a similar percentage of original and MSF faces were fixated, $t(20) = 2.12$, $p = 0.05$. By contrast, fewer faces were fixated in the LSF and HSF conditions, all $ts \geq 3.01$, $ps \leq 0.008$, and in HSF than LSF, $t(20) = 3.26$, $p < 0.008$. For fixated-and-detected trials, a one-factor within-subject ANOVA also showed a main effect of condition, $F(3,60) = 22.82$, $p < 0.001$, $\eta_p^2 = 0.53$. Again, the percentage of trials on which faces were fixated-and-detected was equivalent in the original and MSF conditions, $t(20) = 2.11$, $p = 0.05$, and was higher than for LSF and HSF faces, all $ts \geq 3.08$, $ps \leq 0.008$. In addition, more faces were fixated and detected in LSF than HSF, $t(20) = 3.36$, $p = 0.008$.

The search times for faces are depicted in Figure 2.7 and follow the pattern of response times closely. Thus, a one-factor within-subject ANOVA showed an effect of condition, $F(3,60) = 35.60$, $p < 0.001$, $\eta_p^2 = 0.64$, due to faster search times for

faces in the original condition compared to all other conditions, all $ts \geq 3.77$, $ps \leq 0.008$. In contrast, search times were similar for LSF and MSF faces, $t(20) = 0.88$, $p = 0.39$, and slowest for HSF faces compared to all other conditions, $ts \geq 5.75$, $ps < 0.008$.

Face-absent scenes

For completeness, the mean accuracy and median correct reaction times were also calculated for face-absent scenes. Accuracy was at 97.5% (SE = 1.10). The mean of the median correct reaction times was 4.07 seconds (SE = 0.47).

Discussion

To provide a more direct measurement of the effect of SF on face detection, only the embedded face photographs, but not the scene background, were filtered in Experiment 2. Despite this manipulation, a similar pattern to Experiment 1 was found, whereby the original faces were detected best, both in terms of accuracy and response times, and performance was intermediate for MSF and worst for HSF faces. In addition, the LSF faces were detected as quickly as MSF faces in Experiment 2. However, in contrast to Experiment 1, LSF faces were now also detected more accurately than in the HSF condition. By manipulating only the SF content of the face photographs, detection accuracy for LSF faces was therefore improved. At the same time, these faces could still not match the accuracy of MSF. These findings were confirmed by the eye movement data. This showed identical patterns for the percentage of fixated and fixated-and-detected faces, and the search times also showed the same pattern as observers' response times.

These data indicate that an intermediate level of detail, as provided by MSF, is best for face detection. However, LSF also has a clear advantage over HSF, both in terms of detection speed and accuracy. This indicates that salient low-level cues are of greater importance for face detection and is consistent with the notion of a simple colour-shape template (see Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Hershler & Hochstein, 2005). However, an alternative explanation still exists, as the embedded face photographs in the scenes were filtered for Experiment 2. This includes the area of the face, but also the background of the photograph and its outline. An advantage of manipulating the stimuli in this manner is that the filtering process can affect the boundaries between faces and the background. In LSF, for example, these boundaries are diminished as a result of the image-blur that is introduced by this manipulation (see Figure 2.6 and Figure 2.8). By filtering the faces *and* the background photographs, these effects, which are a natural consequence of the filtering process, were not interfered with by creating more defined but artificial boundaries around the face. As a consequence, however, observers were also given additional information for detection, in the shape of the rectangular outlines of these photographs (see Figure 2.6 and Figure 2.8). In Experiment 2, this provides an additional area in SF that might interact with face detection. This is explored in Experiment 3.

Experiment 3

In Experiment 2, the face *photographs* in the scenes were filtered to provide the different SF conditions. Thus, the target areas (i.e. the faces) include additional, non-face information from the photographs' frames and backgrounds. In Experiment 3, this irrelevant information was removed too, to provide the most direct test yet of

SF for face detection. If the additional information provided by the photographs' frames and backgrounds affected the outcome of Experiment 2, then a different pattern of results should be found here. In turn, if the same pattern of results is found, then this confirms the findings of the experiments reported so far.

Method

Participants

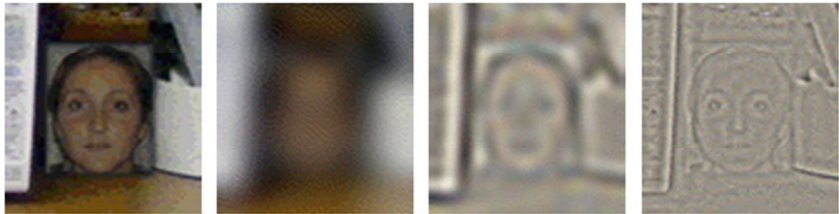
Twenty new undergraduate students (6 males, 14 females) from the University of Kent, with a mean age of 19.5 (SD = 1.5), participated for course credit. All reported normal or corrected-to-normal vision.

Stimuli and procedure

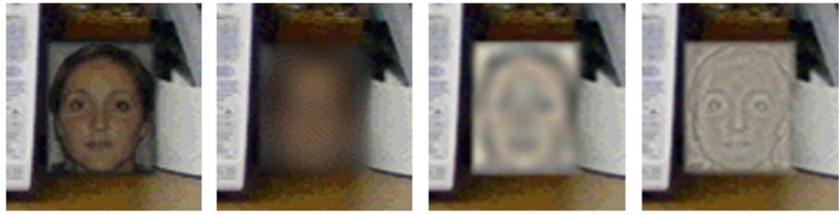
The stimuli and procedure were identical to Experiment 2, except for the following changes. To create face-present scenes, only the target faces were now filtered to display the LSF, MSF or HSF, whereas the surrounding scene content was always displayed in its unfiltered, original format. In addition, the background and outline of the embedded face photographs was removed altogether with a graphics software (Adobe Photoshop CS3) to ensure that this does not interact with face detection. During this process, the edges of the faces were softened slightly, to improve blending into the scene background, by using the "feather" function with a width of 2 pixels. Example stimuli are displayed in Figure 2.8.



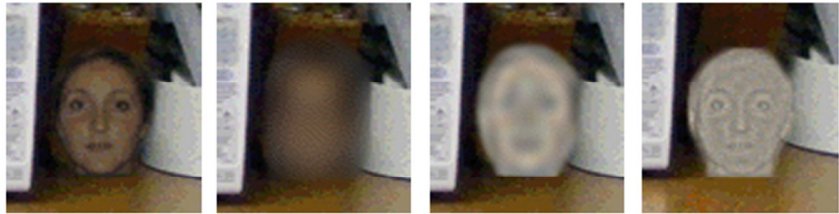
Experiment 1



Experiment 2



Experiment 3



Experiment 4

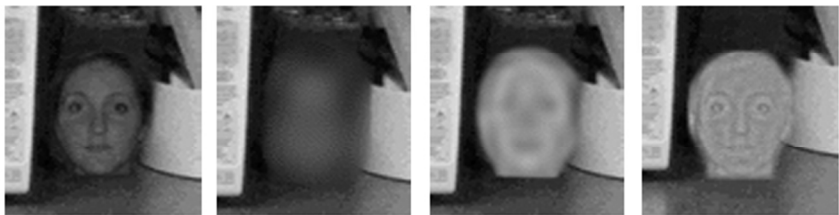


Figure 2.8 A comparison of the stimulus manipulations in Experiments 1 to 4. An example of an original face-present scene is shown at the top. The rows beneath display the face regions of the original, LSF, MSF and HSF conditions (from left to right) for each experiment.

Results

Accuracy and Response times

The data from one participant, whose search and response times were more than two standard deviations from the group mean, was excluded from the analysis. The mean accuracy and median correct response times for the remaining participants are illustrated in Figure 2.9. For the percentage accuracy data, a one-factor within-subject ANOVA showed an effect of condition, $F(3,54) = 34.26$, $p < 0.001$, $\eta_p^2 = 0.67$. T-tests (with *alpha* corrected at $p < 0.008$ for multiple comparisons) showed that accuracy for the original and MSF faces was comparable, $t(18) = 2.39$, $p = 0.03$, and higher than for LSF and HSF faces, all $ts(18) \geq 5.00$, $ps < 0.008$. In addition, accuracy was also higher for LSF than HSF faces, $t(18) = 3.00$, $p < 0.008$.

For the response times, a one-factor within-subject ANOVA also showed an effect of condition, $F(3,54) = 39.95$, $p < 0.001$, $\eta_p^2 = 0.69$, which reflects faster response times in the original condition compared to all other conditions, all $ts \geq 4.04$, $ps \leq 0.008$, while performance for LSF and MSF faces was similar, $t(18) = 1.94$, $p = 0.07$. In contrast, response times to HSF faces were slower than for all other conditions, all $ts \geq 4.74$, $ps < 0.008$.

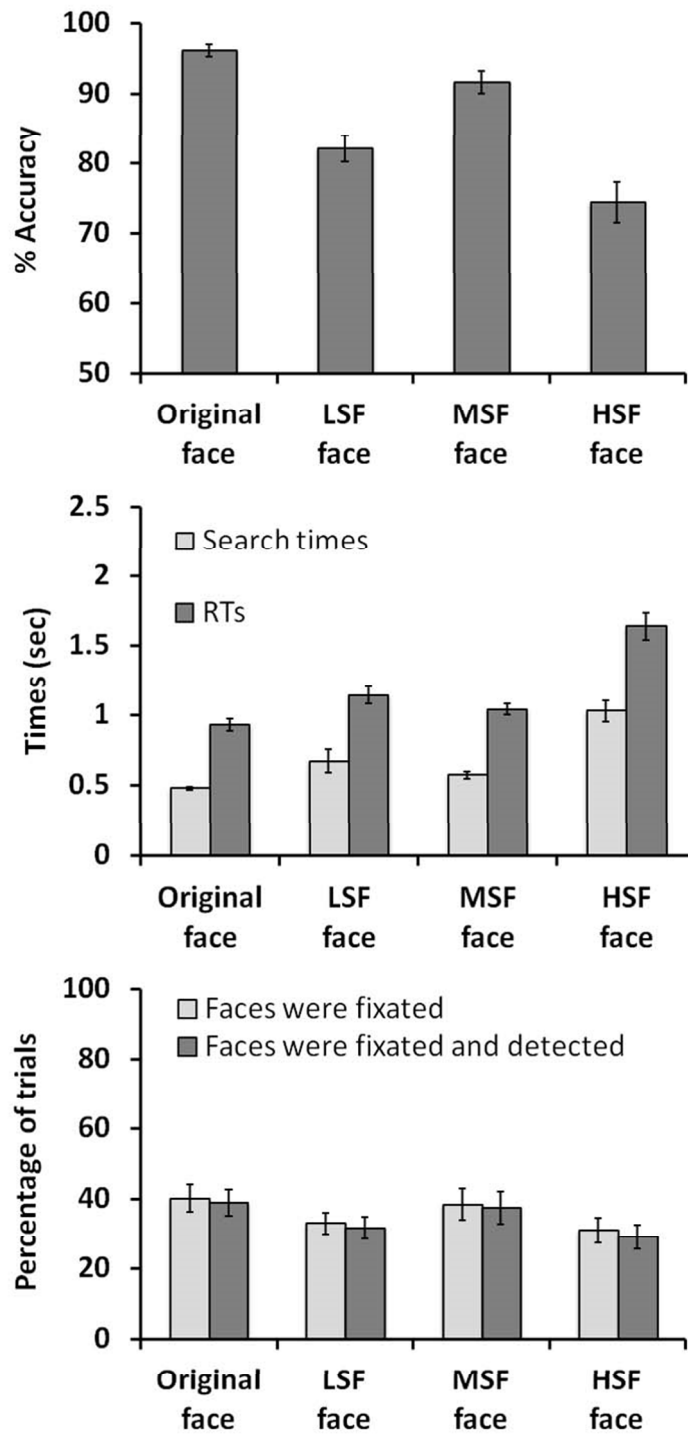


Figure 2.9 Detection performance for face-present scenes in Experiment 3, showing accuracy (top), reaction and search times (middle), and the eye movements to faces (bottom). Vertical bars represent the standard error of the means.

Eye movements

As in the preceding experiments, the mean percentage of trials on which faces were fixated and fixated-and-detected was analysed for all conditions (see Figure 2.9). For both, a one-factor within-subject ANOVA showed a main effect of condition, $F(3,54) = 3.93$, $p < 0.05$, $\eta_p^2 = 0.18$ and $F(3,54) = 4.19$, $p < 0.05$, $\eta_p^2 = 0.19$, respectively. For both measures, Bonferroni t-tests (with *alpha* corrected at $p < 0.008$ for multiple comparisons) found no difference between the original, LSF, and MSF conditions, all $ts(18) \leq 2.36$, $ps \geq 0.03$, except that faces were more likely to be fixated in the original than in the HSF condition on fixated trials, $t(18) = 2.97$, $p < 0.008$, and fixated-and-detected trials, $t(18) = 3.19$, $p < 0.008$.

Search times were analyzed next and are also depicted in Figure 2.9. A one-factor-within-subjects ANOVA of this data showed a main effect of condition, $F(3,54) = 19.93$, $p < 0.001$, $\eta_p^2 = 0.53$, due to faster search times for original than MSF and HSF faces, both $ts \geq 3.68$, $ps \leq 0.008$. In contrast, search times were slowest for HSF faces in comparison with all conditions, all $ts \geq 3.57$, $ps \leq 0.008$. Search times for original and LSF faces, $t(18) = 2.49$, $p = 0.02$, and also for LSF and MSF faces, $t(18) = 1.24$, $p = 0.23$, did not differ.

Face-absent scenes

For completeness, the mean accuracy and median correct reaction times were calculated for face-absent scenes. Accuracy was at 97.1% (SE = 0.36). The mean of the median correct reaction times was 2.93 seconds (SE = 0.22).

Discussion

This experiment replicates the key findings of Experiment 2. Performance was best for the original faces. However, of the filtered conditions, accuracy, response and search times were best for MSF and worst for HSF faces. The LSF faces were detected more frequently than HSF faces, but could only match the detection speed but not the accuracy of MSF. Thus, these findings further support the notion that an intermediate level of detail, such as MSF, is best for face detection. However, LSF faces are detected as quickly as MSF, which suggest that this SF band supports the *fast* detection of faces.

However, a simple explanation for such a low-level advantage might still exist as the original faces and scenes were presented in colour. During filtering, this colour information is preserved in LSF. By contrast, MSF and HSF faces are essentially rendered in greyscale (see Figure 2.8). It is already known that skin-colour tone facilitates face detection, both compared to faces depicted in unnatural colours or greyscale (see Bindemann & Burton, 2009; Lewis & Edmonds, 2003, 2005). This raises the possibility that performance for LSF and MSF faces is driven by different cues. For LSF faces, this might reflect the available colour information rather than SF content. In contrast, the detection of MSF faces might be supported by the intermediate SF content of these stimuli rather than colour cues. To explore these possibilities, the next experiment explored the detection of these faces in greyscale scenes.

Experiment 4

In contrast to Experiment 3, which compared detection performance of SF faces in their natural colour, the current experiment presented the faces and scenes in greyscale. If colour information drives the detection advantage of LSF over HSF faces, then this effect should disappear in the current conditions. In contrast, if this detection advantage is determined by the LSF information that is preserved in these faces, irrespective of their colour content, then the same pattern as in the preceding experiments should be found.

Method

Participants, stimuli and procedure

Twenty two undergraduate students (4 males, 18 females) from the University of Kent, with a mean age of 19.3 (SD = 1.2), participated in the study in exchange for course credits. All participants reported normal or corrected-to-normal vision. The stimuli and procedure were identical to Experiment 3 except that all stimuli were transformed into greyscale scale using the standard function (the grayscale image mode) of a graphics software (Adobe Photoshop CS3). Example stimuli are depicted in Figure 2.8.

Results

Accuracy and response times

The mean accuracy and median correct response times are illustrated in Figure 2.10. For the percentage accuracy data, a one-factor within-subject ANOVA revealed an effect of condition, $F(3,63) = 43.80, p < 0.001, \eta_p^2 = 0.68$. Bonferroni t-tests (with *alpha* corrected at $p < 0.008$ for multiple comparisons) showed that accuracy for original and MSF faces was comparable, $t(21) = 1.99, p = 0.06$, and was better than for LSF and HSF faces, all $ts > 7.02, ps < 0.001$. In addition, accuracy for LSF and HSF faces did not differ, $t(21) = 0.51, p = 0.61$.

For the response times, a one-factor within-subject ANOVA also showed an effect of condition, $F(3,63) = 34.15, p < 0.001, \eta_p^2 = 0.62$. Bonferroni t-tests again revealed similar response times for the original and MSF faces, $t(21) = 2.63, p = 0.02$, and for LSF and MSF faces, $t(21) = 2.18, p = 0.04$, but quicker response times for the original than LSF faces, $t(21) = 3.39, p = 0.003$. In addition, response times were slower for HSF faces compared to all other conditions, all $ts \geq 4.90, ps < 0.001$.

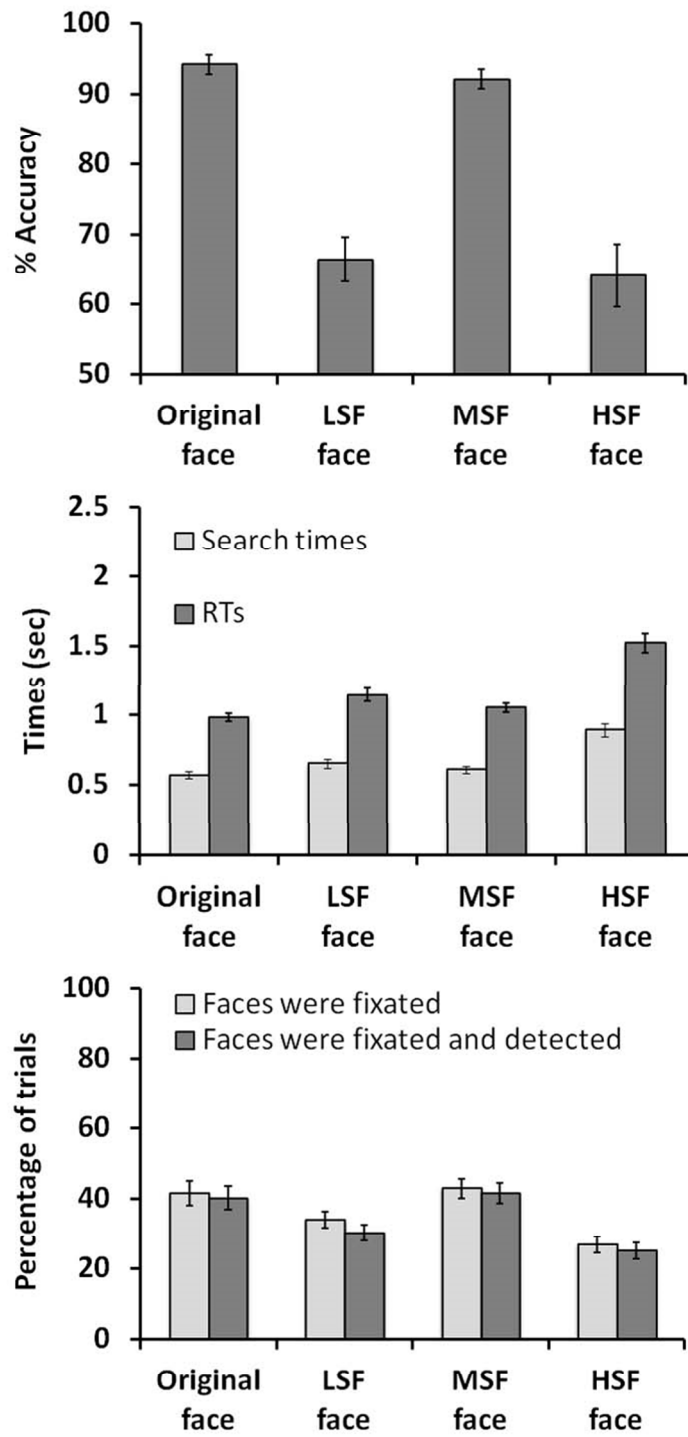


Figure 2.10 Detection performance for face-present scenes in Experiment 4, showing accuracy (top), reaction and search times (middle), and the eye movements to faces (bottom). Vertical bars represent the standard error of the means.

Eye movements

As in the preceding experiments, the mean percentage of trials on which faces were fixated and fixated-and-detected was analysed for all conditions (see Figure 2.10). For the fixated trials, a one-factor within-subject ANOVA showed a main effect of condition, $F(3,63) = 12.31, p < 0.001, \eta_p^2 = 0.37$. The percentage of trials on which the original faces were fixated was comparable to LSF and MSF faces, both $ts \leq 2.10, ps \geq 0.05$ (with *alpha* corrected at $p < 0.008$ for multiple comparisons), and was higher for original and MSF than HSF faces, both $ts \geq 4.42, ps < 0.008$. In addition, faces were also more likely to be fixated in MSF than LSF, $t(21) = 3.84, p < 0.008$, whereas LSF and HSF faces did not differ, $t(21) = 2.31, p = 0.03$.

A similar pattern was obtained on fixated-and-detected trials, for which ANOVA also revealed a main effect of condition, $F(3,63) = 11.91, p < 0.001, \eta_p^2 = 0.36$. The percentage of fixated-and-detected trials was comparable for the original and MSF faces, $t(21) = 0.48, p = 0.64$, and higher than for HSF faces, both $ts \geq 4.53, ps < 0.008$. In addition, this percentage was also higher for MSF than LSF faces, $t(21) = 3.83, p < 0.008$, but similar for original and LSF faces, $t(21) = 2.62, p = 0.02$, and for LSF and HSF faces, $t(21) = 1.54, p = 0.14$.

Search times for fixated-and-detected trials were analyzed next and are also depicted in Figure 2.10. A one-factor within-subjects ANOVA of this data showed a main effect of condition, $F(3,63) = 16.73, p < 0.001, \eta_p^2 = 0.44$. Search times were comparable for original, LSF and MSF faces, all $ts \leq 1.76, ps \geq 0.09$, and were slowest for HSF faces compared to all other conditions, all $ts \geq 3.90, ps \leq 0.008$.

Face-absent scenes

For completeness, the mean accuracy and median correct reaction times were calculated for face-absent scenes. Accuracy was at 97.9% (SE = 0.30). The mean of the median correct reaction times was 2.47 seconds (SE = 0.16).

Discussion

Despite the removal of colour information, this experiment replicates the key findings of the preceding experiments. As in Experiments 1 to 3, detection was fastest and most accurate for original and MSF faces, and worst for faces displayed in HSF. In contrast, detection of LSF faces was as fast as MSF, both in terms of response and search times. These findings suggest that the speed advantage of LSF faces is not only determined by the colour content of these stimuli, which is not preserved in MSF and HSF, but must be related more directly to the SF content. At the same time, the detection of LSF faces was less accurate than for original and MSF faces. This effect was such that, in comparison to Experiment 2 and 3, detection accuracy was now at the same level for LSF and HSF faces. This contrast indicates that colour information is still beneficial for face detection and is consistent with the notion that this process might be driven by a skin-coloured face-shape template (see Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Hershler & Hochstein, 2005). At the same time, these findings are also consistent with the notion of a ‘quick and dirty’ processing strategy that is driven by low-level cues (Crouzet & Thorpe, 2011), by demonstrating that grayscale LSF content is sufficient for fast (but not always accurate) face detection.

Before reaching this conclusion, an alternative explanation still needs to be eliminated. In Experiment 1, the faces and scene background were filtered together to produce the SF conditions. Consequently, it was unclear whether differences between conditions reflected face or scene processing. Experiments 2 to 4 therefore explored detection by presenting SF faces in unfiltered background images. However, this manipulation might also bias results. For example, it is conceivable that the regions in these scenes that have been rendered in different SF attract observers' attention, rather than the face content *per se*. Indeed, observers could explicitly choose to adopt such a strategy if the filtered regions contrast somehow with the image quality of the surrounding scene. In line with this reasoning, it has already been shown that small, blurred regions within visual displays can attract observers' eye movements (Smith & Tadmor, 2013). If a similar effect is found here, then the fast detection of MSF and LSF faces might reflect the contrast between the blurred regions and the surrounding high-resolution scene context rather than the underlying facial information. This possibility is explored in a final experiment.

Experiment 5

Experiment 5 investigates whether the pattern of the preceding experiments reflects face detection or observers' sensitivity to patches of different SF content when these are embedded within high-resolution scenes. To explore this possibility, this experiment reverts to the colour scenes of Experiment 3 but provides additional control conditions as face-absent scenes. In these scenes, the locations that correspond to the faces in their face-present counterparts are selectively rendered in LSF and HSF (see Figure 2.11). If the detection pattern of the preceding experiments reflects the saliency of SF patches within high-resolution scenes, then these regions should attract

observers eye movements regardless of whether these coincide with the location of a face (in face-present scenes) or not (in face-absent scenes). In this case, observers should fixate LSF regions faster than HSF patches. In turn, if such an effect is not found, then this will confirm that the findings of the preceding experiments reflect the role of spatial frequency in face detection.



Figure 2.11 An example of a face-absent scene in Experiment 5, depicting the original (top), LSF (middle), and HSF conditions (bottom). SF is manipulated in small patches, which correspond to the size and location of face photographs in the corresponding face-present scenes.

Method

Participants

Twenty-three new undergraduate students (5 male, 18 female) from the University of Kent, with a mean age of 21.6 years ($SD = 4.2$), participated in exchange of credits. All participants reported normal or corrected-to-normal vision.

Stimuli and procedure

The stimuli and procedure were identical to Experiment 3, which assessed face detection with colour scenes, except for the following changes. In each of the original face-absent scenes, small patches were filtered to preserve only LSF and HSF information. The location and size of these patches matched that of the faces in the corresponding face-present scenes. This resulted in a total of 360 face-absent scenes, comprising 120 images in the original, LSF and HSF conditions, and 360 face-present scenes in the same SF conditions.

To ensure that observers cannot predict the location of the SF patches in the face-present or face-absent scenes, the same scene backgrounds were not presented repeatedly to any of the participants. In contrast, the original 120 indoor photographs were separated into 60 scenes for each of the face-present and face-absent conditions, comprising 20 images in the original, LSF and HSF conditions. However, over the course of the experiment, the presentation of scenes was counterbalanced across participants and conditions. As in previous experiments, all stimuli were presented in a randomly-intermixed order and participants were asked to determine the presence or absence of faces as quickly and as accurately as possible.

Results

Accuracy and response times

The data from one participant, whose search and response times were more than two standard deviations from the group mean, was excluded from the analysis. The mean accuracy and median correct response times for the remaining participants are illustrated in Figure 2.12 and are analysed first for face-present trials. For accuracy, a one-factor within-subject ANOVA revealed an effect of condition, $F(2,42) = 31.26, p < 0.001, \eta_p^2 = 0.60$. A series of t-tests (with *alpha* corrected at $p < 0.017$ for three comparisons, i.e. $p = 0.05/3$) showed that observers were more accurate in detecting faces in the original than the LSF and HSF conditions, respectively, both $ts \geq 5.50, ps < 0.017$, whereas detection accuracy was more similar for LSF and HSF faces, $t(21) = 2.52, p = 0.02$. Response times revealed a similar pattern. ANOVA showed a main effect of condition, $F(2,42) = 30.07, p < 0.001, \eta_p^2 = 0.59$, which reflects faster detection of original than LSF and HSF faces, all $ts \geq 3.69, ps \leq 0.017$. In addition, LSF faces were detected faster than HSF faces, $t(21) = 4.40, p < 0.017$.

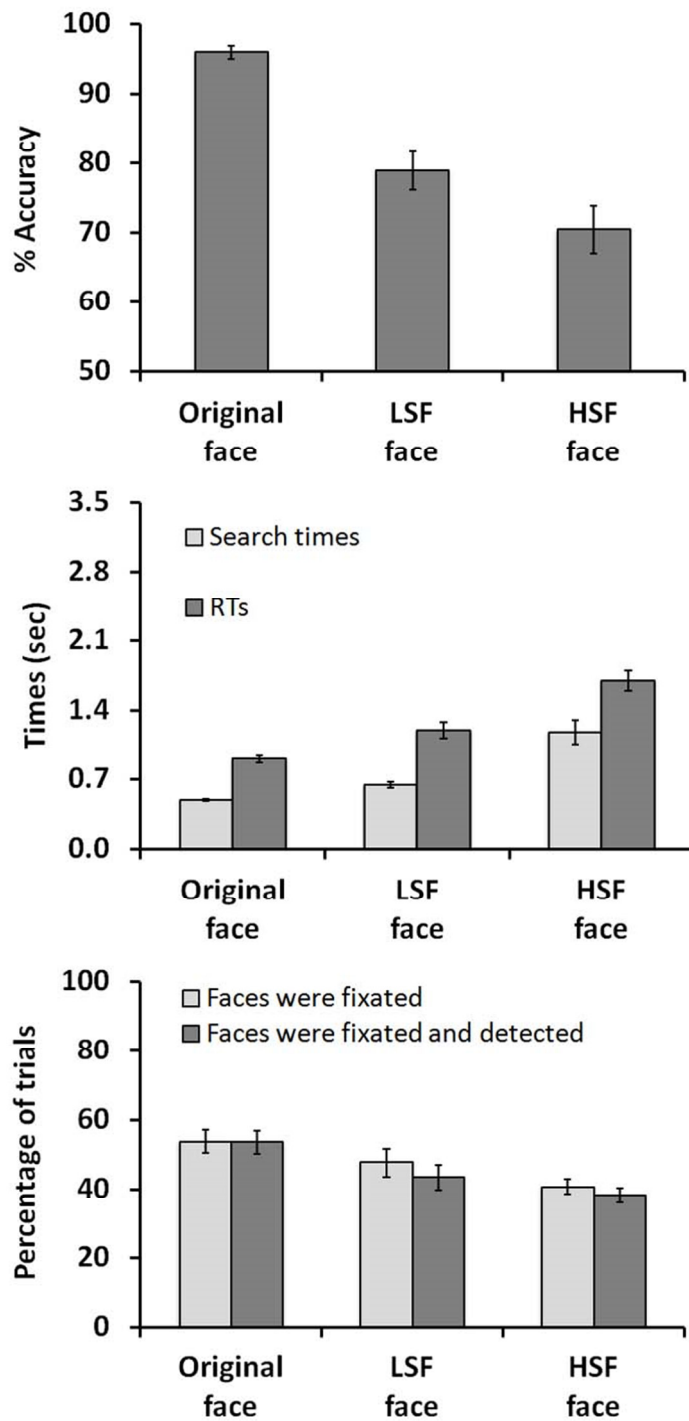


Figure 2.12 Detection performance for face-present scenes in Experiment 5, showing accuracy (top), reaction and search times (middle), and the eye movements to faces (bottom). Vertical bars represent the standard error of the means.

Eye movements

As in the preceding experiments, eye movements were processed also to assess face detection across conditions. Three measures were assessed corresponding to mean of the percentage of trials on which faces were fixated or fixated-and-detected in each condition, and the search times to first fixate a face in a scene (see Figure 2.12).

For all fixated trials, a one-factor within-subject ANOVA showed an effect of condition, $F(2,42) = 8.82$, $p < 0.01$, $\eta_p^2 = 0.30$. Bonferroni t-tests (with *alpha* corrected at $p < 0.017$ for multiple comparisons) revealed that faces were more likely to be fixated in the original than the HSF condition, $t(21) = 4.67$, $p < 0.017$, whereas the percentage of fixations for original and LSF faces, $t(21) = 1.96$, $p = 0.06$, and for LSF and HSF faces, $t(21) = 2.05$, $p = 0.06$, did not differ. For the fixated-and-detected trials, an effect of condition was also found, $F(2,42) = 13.34$, $p < 0.001$, $\eta_p^2 = 0.39$, which reflects a higher percentage score for the original than LSF and HSF faces, both $ts \geq 3.35$, $ps \leq 0.017$. The LSF and HSF conditions did not differ, $t(21) = 1.64$, $p = 0.12$. Finally, the search times also showed an effect of condition, $F(2,42) = 27.59$, $p < 0.001$, $\eta_p^2 = 0.57$. Faces were fixated more quickly in the original than the LSF and HSF conditions, both $ts \geq 5.64$, $ps < 0.017$, and also in the LSF than the HSF condition, $t(21) = 3.52$, $p < 0.017$.

Face-absent scenes

Observers' responses to face-absent scenes were also analysed. Accuracy, response times and eye movement data for these scenes are depicted in Figure 2.13. Generally, performance across the face-absent condition was very similar.

Nonetheless, a main effect of condition was found for accuracy, $F(2,42) = 5.00$, $p < 0.05$, $\eta_p^2 = 0.19$. This was assessed further with Bonferroni t-tests (with *alpha* corrected at $p < 0.017$ for multiple comparisons), which showed that more absent responses were made in the HSF condition compared to the LSF and original conditions, both $ts \geq 3.05$, $ps \leq 0.017$. However, the original and LSF condition did not differ, $t(21) = 1.14$, $p = 0.27$. Response times were also similar across conditions and an effect of condition was not found, $F(2,42) = 1.24$, $p = 0.30$, $\eta_p^2 = 0.06$.

The analysis of main interest here concerned the extent to which the SF regions were fixated, and how quickly this happened, in the face-absent scenes. This should determine whether the results for face-present scenes reflect face detection processes or are driven simply by the saliency of SF patches within high-resolution scenes. The mean percentage of trials on which these SF regions were fixated in face-absent scenes shows an effect of condition, $F(2,42) = 9.92$, $p < 0.001$, $\eta_p^2 = 0.32$. This arises because LSF patches were fixated more frequently than HSF patches, $t(21) = 2.97$, $p < 0.017$. In addition, fixations to this region were also higher in the LSF than the original condition, $t(21) = 4.11$, $p < 0.017$. By contrast, no difference was found between the original and HSF condition, $t(21) = 1.37$, $p = 0.17$. Overall, however, the percentage of trials on which all of these regions were fixated was small compared to face-present scenes (c.f. Figures 2.12 and 2.13). Moreover, the speed with which these regions were first fixated (search times) did not differ across the original, LSF and HSF conditions, $F(2,42) = 1.19$, $p = 0.31$, $\eta_p^2 = 0.05$.

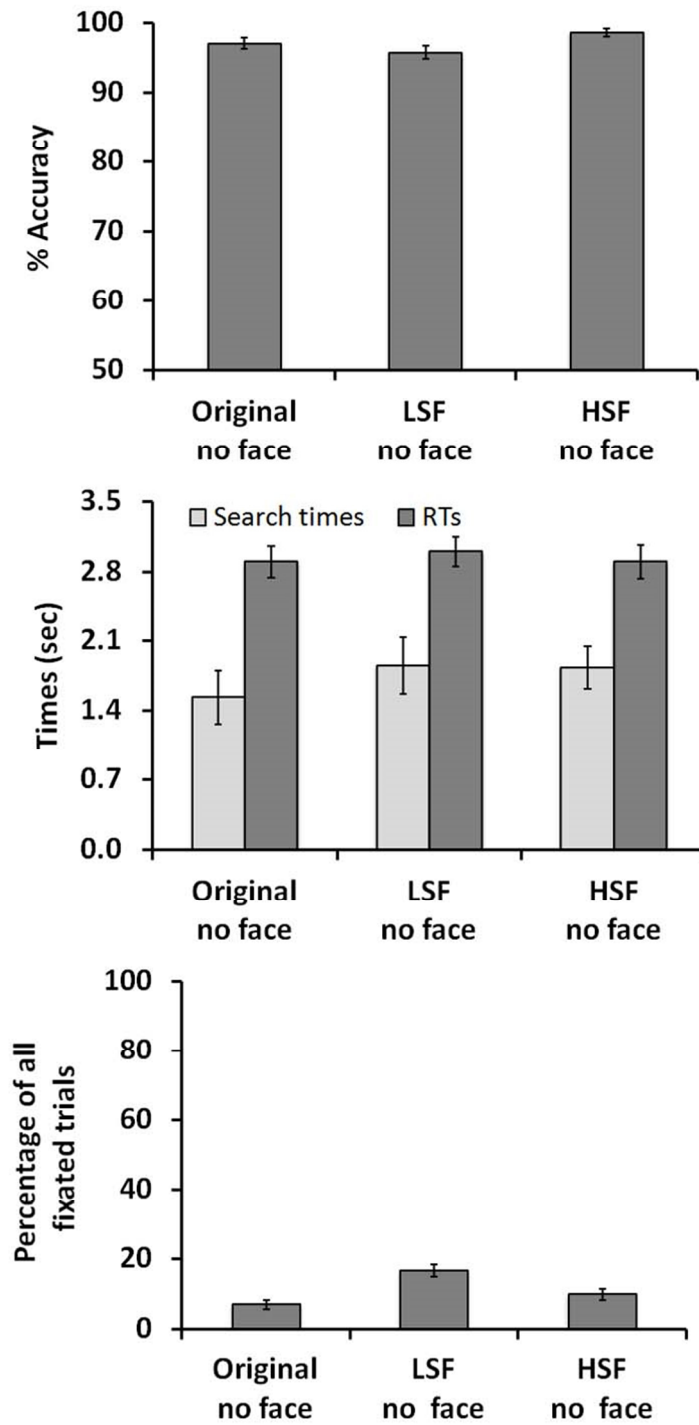


Figure 2.13 Detection performance for face-absent scenes in Experiment 5, showing accuracy (top), reaction times (middle), and the percentage of trials on which the SF patches' locations were fixated (bottom). Vertical bars represent the standard error of the means.

Discussion

This experiment examined whether the pattern of the preceding experiments reflects face detection processes or observers' sensitivity to patches of different SF content when these are embedded within high-resolution scenes. This could reflect a strategy whereby observers scan scenes for regions that are visually different from the surrounding high-resolution context. For this purpose, this experiment included additional face-absent conditions in which small regions within these scenes that correspond to the location and size of faces in the face-present scenes were filtered to retain only LSF or HSF information.

Consistent with all of the preceding experiments, detection was best for original faces, followed by LSF and HSF faces. In face-absent scenes, LSF patches were fixated more often than HSF patches. On its own, this might suggest that low-level artefacts, such as different SF regions within a scene, could contribute to the face detection effects in the preceding experiments. However, the percentage of trials on which these regions were fixated was low compared to face-present scenes. More importantly, LSF patches were not classified more accurately or faster, or were fixated quicker, than HSF scenes. Taken together, these findings indicate that the results of the preceding experiments are not simply an artefact of the experimental manipulations but reflect the role of specific spatial frequencies for face detection.

General Discussion

This study examined how spatial frequency affects the detection of faces in natural scenes. In Experiment 1, a clear effect of SF was found, whereby faces were

detected fastest when scenes were presented in their original condition, but were also detected faster in MSF and LSF than HSF displays. Detection accuracy was also highest in the original condition, reduced in MSF, and considerably lower for HSF displays. However, in contrast to reaction times and search times, which indicated better detection for LSF than HSF faces, accuracy across these two conditions was more evenly matched. In Experiment 1, it is possible that the scene background, which was also rendered in the different spatial frequencies, might have contributed to these effects. In subsequent experiments, only the face photographs within the scenes (Experiment 2) or the face regions within these photographs (Experiment 3) were therefore rendered in different spatial frequencies, while the scene background remained intact. In both experiments, detection accuracy for original and MSF faces was best, followed by LSF and HSF faces, respectively. The original faces were also located fastest and HSF faces slowest of all, but performance for MSF and LSF faces was comparable. A consistent pattern emerges from these experiments, whereby face detection performance is reliably best when all SF are preserved (in the original condition) and worst for HSF, both in terms of detection speed and accuracy. By contrast, MSF and LSF faces are detected with similar speed but MSF faces are detected more accurately than LSF faces. These findings suggest that LSF support the fast detection of faces. Occasionally, however, this SF band also provides insufficient detail to locate a face at all.

A further experiment explored whether the fast detection of LSF faces reflects the colour information in these stimuli, which was not preserved in the MSF and HSF conditions in Experiments 1 to 3. For this purpose, all stimuli were rendered in greyscale in Experiment 4. Despite the removal of this information, Experiment 4

replicated the key findings of the preceding experiments, by demonstrating similar detection speeds for LSF and MSF faces, and worst performance for HSF. This is an important finding that indicates that the fast detection of LSF faces is not simply driven by colour information. Instead these findings suggest that, despite the hugely impoverished facial representations that these stimuli provide (see Figures 2.2, 2.6 and 2.8), LSF faces must also contain some basic structural information that is sufficient to support fast face detection.

A final experiment then investigated whether the current findings could reflect a low-level image artefact, whereby observers search scenes for SF regions with the original scenes rather than looking for faces. To explore this possibility, SF was manipulated selectively in small patches of the face-absent scenes, which corresponded to the size and location of the face photographs in face-present scenes. In this experiment, detection was best once again for original faces, followed by the LSF and HSF conditions (an MSF condition was not included in Experiment 5). In contrast, the SF patches of face-absent scenes were fixated much less frequently than faces, and LSF patches were not classified more accurately, faster, or were fixated quicker than HSF patches. These findings therefore suggest that the results of the preceding experiments cannot be explained by a low-level image artefact, such that observers simply search for SF regions in scenes. Instead, these findings suggest that the effects of the preceding experiments reflect the removal of SF information from *faces*.

A possible explanation for the differences in detection performance could be that SF faces are more difficult to detect because they do, in fact, look less like faces. The current data do not speak to this directly, but observers were informed about the

conditions prior to the experiments. In addition, the current data are also consistent with the notion that our visual system is developed to detect faces in the visual periphery, based on LSF information (Johnson, 2005). Moreover, the rapid detection of LSF and MSF faces supports the idea of a 'quick and dirty strategy' for face detection via low-level cue information (Crouzet & Thorpe, 2011). Considering the highly impoverished nature of the LSF faces, this might be driven by a simple template, such as a face-shaped oval. This notion converges with other experiments that have shown that face detection proceeds unhindered as long as a round face-shape is preserved (Hershler & Hochstein, 2005). In those experiments, this was found to be the case even when facial features, such as the eyes, nose and mouth, were removed. This also converges with the current findings, which suggest that facial features, as captured by HSF but not clearly visible in LSF, are not of primary importance for face detection. At the same time, some of those features, such as the eyes, clearly help face detection when overall face-shape is compromised (see Burton & Bindemann, 2009). Thus, face detection appears to utilize multiple sources of information, including the detail of HSF. However, the current results indicate that the facial aspects captured by LSF specifically support the *fast* detection of faces.

The fast detection of faces might be governed mainly by the magnocellular brain pathway. This channel reportedly supports LSF processing (Bullier, 2001; Livingstone & Hubel, 1988) and also appears to be suited best to the detection of faces in the visual periphery (Awasthi, Friedman, & Williams, 2011a, 2011b). By contrast, higher spatial frequencies, which code finer visual details of faces, are held to be processed by a slower, ventral stream in the human cortex, via the fovea-sensitive parvocellular channel (Bullier, 2001; Livingstone & Hubel, 1988; Lynch,

Silveira, Perry, & Merigan, 1992). While the magnocellular brain pathway might support the fast detection of LSF faces from LSF, in complex natural scenes, such as the stimuli of the current task, the parvocellular channel might help to maximise performance in the original and MSF conditions, or when LSF and MSF information is sub-optimal or unavailable (e.g. in HSF displays). In line with this reasoning, it has already been shown that early stages of face processes, as measured via the N170 event-related potential, operate best when LSF and HSF information are both available (Halit, de Haan, Schyns, & Johnson, 2006). The current experiments suggest as much, by consistently showing that performance is best in the original face condition, in which all SF bands are preserved.

In such a framework, the SF information processed by these two streams essentially performs the same purpose in parallel, but proceeds at different speeds. However, it might also be possible that LSF and HSF have different roles, whereby the former is used to quickly identify possible face candidates and the latter then helps to confirm that a looked-at stimulus is, in fact, a face (for similar suggestions, see Bindemann & Lewis, 2013). Despite the appeal of such a framework, several aspects of the current data speak against such a two-stage process. Firstly, in such a framework, one might expect that faces are occasionally fixated but not detected. This might be the case particularly under LSF, which seem to provide very salient face cues but little visual detail. This might be sufficient to locate stimuli that are possible face candidates but could on occasion also be insufficient to confirm that a looked-at stimulus is a face.

To explore this in the analysis of eye movements, the percentage of trials on which faces were fixated and the percentage of trials on which they were fixated-and-

detected were analysed. These two measures consistently returned very similar values and identical patterns across conditions, which suggests that it is unlikely that faces were detected as possible candidates in the current experiment but not confirmed as such. To explore this more directly, a further analysis is conducted here to directly assess the percentage of trials on which faces were *fixated-but-not-detected* across conditions. This data is given in Figure 2.14 for Experiments 2 to 4 and shows that the percentage of these cases was generally low and similar across conditions. In line with these observations, a one-factor ANOVA failed to find an effect of condition for the percentage of fixated-but-not-detected faces in all three experiments, all $F_s \leq 2.17$, $p_s \geq 0.15$, $\eta_p^2 \leq 0.19$.

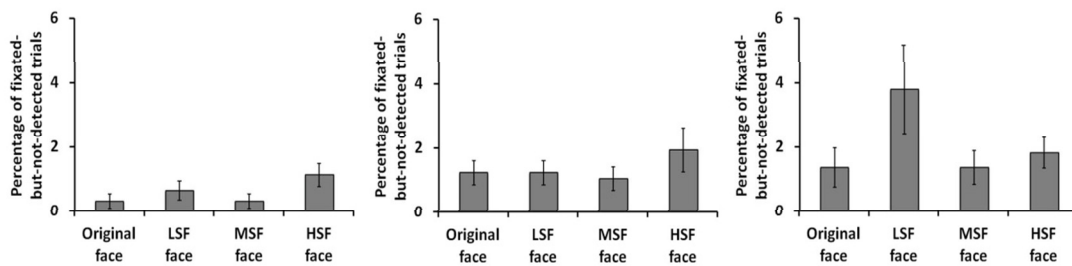


Figure 2.14 The percentage of fixated-but-not-detected trials in Experiment 2 (left panel), Experiment 3 (middle panel), and Experiment 4 (right panel). Vertical bars represent the standard error of the means.

Secondly, in a two-stage framework one might also expect that search times (observers' eye movements) and response times reflect different aspects of face detection, whereby the former measures search for likely face candidates (and is supported by LSF) and the latter the (search and) decision that a fixated stimulus is, in fact, a face (supported by HSF). Contrary to this notion, however, these two measures consistently returned the same pattern across conditions in the current experiments, which suggests that both reflect the same underlying process. This issue was explored further by subtracting search from response times. The difference between these

scores might provide a more direct reflection of the *decision* time that is required to confirm that a looked-at stimulus really is a face. If this process is separable from search for likely face candidates, and depends more on HSF information, then this should show an advantage for those spatial frequencies. Again, however, ANOVA failed to find an effect of condition in Experiment 2 and 3, both $F_s \leq 3.09$, $p_s \geq 0.08$, $\eta_p^2 \leq 0.16$ (see Figure 2.15). Experiment 4 revealed an effect of condition, $F(3,63) = 6.82$, $p = 0.002$, $\eta_p^2 = 0.245$, but, contrary to prediction, this reflects slower decision times for HSF than the original and MSF faces, both $t_s \geq 3.45$, $p_s \leq 0.002$ (and no other comparisons were significant, all $t_s \leq 1.27$, $p_s \geq 0.22$). Taken together, these data suggest that face detection does not reflect a two-stage process, of the initial search for likely face candidates and a subsequent decision stage. Instead, these processes appear to be inseparable in the current paradigm. However, more direct investigations of this theory are still clearly needed.

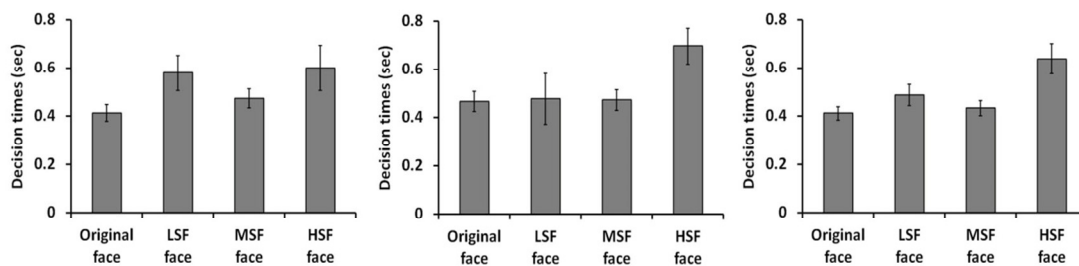


Figure 2.15 Decision times in Experiment 2 (left panel), Experiment 3 (middle panel), and Experiment 4 (right panel). Vertical bars represent the standard error of the means.

At this stage, the finding that LSF facilitates quick but not accurate detection still raises further questions. It is already known that LSF is particularly useful for locating faces quickly in the periphery (Awasthi, Friedman, & Williams, 2011a), but the current experiments also show that MSF faces are detected as fast as LSF faces and with higher accuracy. While the current study applied SF based on the average of

face sizes (59 (H) x 47 (W) pixels), the size of individual faces ranged from 36 x 27 to 139 x 115 pixels. This raises the question of whether face size might have affected the contribution of SF in the current experiments and could have obscured the advantage of LSF.

To begin to explore this *a posteriori*, the response times and search times for each face condition were re-calculated as a function of face sizes. While face sizes ranged from 575-10150 pixels, only very few faces measured more than 3000 pixels. The stimuli were therefore divided into three large non-overlapping face categories of small, medium and large size, reflecting sizes of less than 1500 pixels, 1501-3000 pixels, and 3001-10150 pixels, respectively. This data is illustrated in Figure 2.16. A series of 4 (face conditions: original, LSF, MSF, and HSF) x 3 (face sizes: original, medium, and large) ANOVAs of response times and search times were conducted on this data for Experiments 2 to 4.

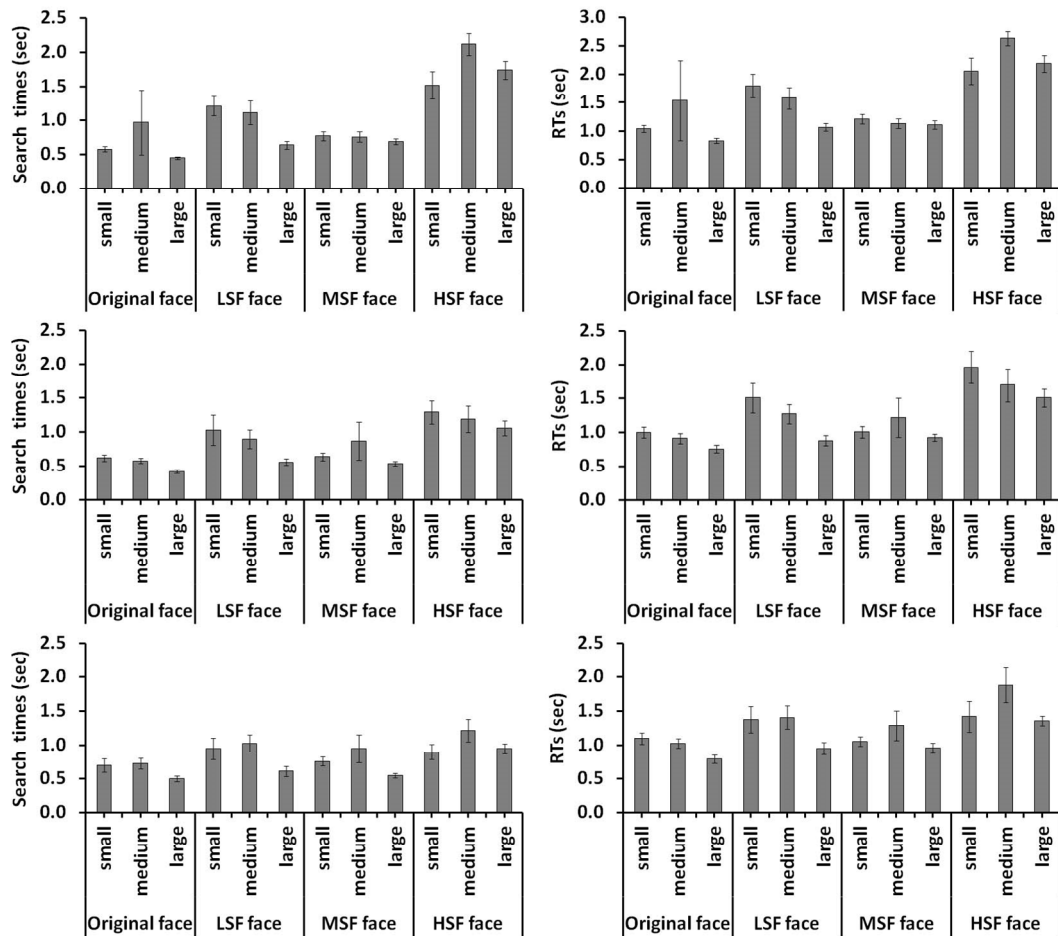


Figure 2.16 Search times (left) and response times (right) for each face condition in Experiment 2 (top panel), 3 (middle panel) and 4 (bottom panel), grouped by face sizes. Vertical bars represent the standard error of the means.

Generally, response times across the three experiments were faster for original faces, followed by MSF, LSF and HSF, particularly when the size of the faces was large. For Experiment 2, 3 and 4, the analysis of response times showed main effects of face condition (Experiment 1: $F(3,45) = 16.30, p < 0.001, \eta_p^2 = 0.52$; Experiment 2: $F(3,39) = 25.16, p < 0.001, \eta_p^2 = 0.67$; Experiment 3: $F(3,39) = 8.44, p < 0.001, \eta_p^2 = 0.39$), due to slower response times in the HSF conditions, all $t_s \geq 4.49, p_s < 0.001$ (Bonferroni t-tests with α corrected at $p < 0.008$, i.e. $p = 0.05/6$), and main effects of face size (Experiment 1: $F(2,30) = 4.02, p < 0.05, \eta_p^2 = 0.21$; Experiment 2: $F(2,26) = 4.24, p < 0.05, \eta_p^2 = 0.25$; Experiment 3: $F(2,26) = 7.10, p < 0.01, \eta_p^2 =$

0.35), due to quicker response time in the large face size condition, all $ts \geq 2.41$, $ps \leq 0.016$ (with *alpha* corrected at $p < 0.017$, i.e. $p = 0.05/3$). For all three experiments, the interactions between both factors were not significant (Experiment 1: $F(6,90) = 1.11$, $p = 0.37$, $\eta_p^2 = 0.07$; Experiment 2: $F(6,78) = 0.89$, $p = 0.51$, $\eta_p^2 = 0.06$; Experiment 3: $F(6,78) = 0.78$, $p = 0.59$, $\eta_p^2 = 0.06$).

A similar pattern was obtained for search times, which were generally quicker for original faces, followed by MSF, LSF and HSF faces, and also when faces were large. A series of 4 x 3 ANOVAs of this data also showed main effects of face condition (Experiment 1: $F(3,45) = 30.18$, $p < 0.001$, $\eta_p^2 = 0.67$; Experiment 2: $F(3,39) = 17.06$, $p < 0.001$, $\eta_p^2 = 0.57$; Experiment 3: $F(3,39) = 5.96$, $p < 0.01$, $\eta_p^2 = 0.31$), due to slower search times in the HSF conditions, all $ts \geq 5.93$, $ps < 0.001$ (Bonferroni t-tests with *alpha* corrected at $p < 0.008$ for multiple comparisons), and main effects of face size (Experiment 1: $F(2,30) = 5.67$, $p < 0.01$, $\eta_p^2 = 0.27$; Experiment 2: $F(2,26) = 3.84$, $p < 0.05$, $\eta_p^2 = 0.23$; Experiment 3: $F(2,26) = 7.22$, $p < 0.01$, $\eta_p^2 = 0.36$), due to quicker search time in the large face size condition, all $ts \geq 2.54$, $ps \leq 0.014$ (with *alpha* corrected at $p < 0.017$ for multiple comparisons). None of the interactions were significant (Experiment 1: $F(6,90) = 1.20$, $p = 0.32$, $\eta_p^2 = 0.07$; Experiment 2: $F(6,78) = 0.66$, $p = 0.68$, $\eta_p^2 = 0.05$; Experiment 3: $F(6,78) = 0.75$, $p = 0.62$, $\eta_p^2 = 0.05$). Overall, these data therefore suggest larger faces are detected better than smaller faces. However, the contribution of each range of spatial frequencies is not influenced by face sizes, confirming the advantage of MSF over LSF for face detection.

Whereas variation in face size generally cannot explain why LSF faces are detected as quickly but less accurately than MSF faces, it is also possible that this

might reflect particular stimuli, in which LSF faces are more likely to be missed. To explore this possibility, the trials on which faces were fixated but not detected were also analysed for Experiments 1 to 4 (see Figure 2.14). This analysis shows that such fixated but not detected cases generally occurred with equal frequency across the SF conditions. This suggests that the low accuracy for LSF faces is not caused by the limitation of any single stimulus to provide visual information for detection. However, other explanations for the effect of LSF on face detection remain of course possible.

One of these explanations might relate to the eccentricity of faces within the scene stimuli. LSF is particularly useful for processing faces in the visual periphery (Awasthi, Friedman, & Williams, 2011a). Whereas the current study shows that MSF is most useful when faces occur at a range of sizes and locations. For more direct investigations, a clearer advantage for LSF faces might therefore be found when faces occur at extreme eccentricity in scenes. Similarly, it is possible that a higher cut-off for filtering LSF faces, such as 8 cycles/face (Awasthi, Friedman, & Williams, 2011a, 2011b, Awasthi, Sowman, Friedman, & Williams, 2013; Halit, de Haan, Schyns, & Johnson, 2006) might enhance detection accuracy for this class of stimuli as a function of an improvement of contrast sensitivity (see Campbell & Robson, 1968).

In the meantime, the question also remains of what information is preserved in LSF faces that could drive the fast detection speed for these stimuli. Colour information facilitates detection, but only when this is tied to face-shape (Bindemann & Burton, 2009), and could therefore reflect one of the diagnostic characteristics of an LSF template. However, Experiment 4 also showed that detection of LSF faces remains as fast as for MSF faces when colour information is removed. Thus, LSF faces carry additional information, beyond colour, to facilitate their fast detection.

One possibility is that this reflects gross holistic or configural information about faces, such as simple shading cues, from features such as the eyes, that might be preserved in these stimuli (see Burton & Bindemann, 2009). Alternatively, this might reflect some general dimensions, such as the height and width of faces, and the ratio of these measures. This is explored in the next chapter.

Chapter 3:

The Shape of the Face Template:
Geometric Distortions of Faces and
Their Detection in Natural Scenes.

Introduction

The experiments of the preceding chapter consistently showed that very simple visual structures containing salient colour and shape cues sufficed for rapid detection. In contrast, fine visual information, such as the featural detail of the eyes, nose and mouth delayed detection. This indicates that this process might rely on a “quick and dirty” processing strategy that utilizes salient visual cues to locate likely face candidates (Crouzet & Thorpe, 2011). One possibility for such a strategy could be based on a simple skin-coloured face-shaped template. This idea is based on the finding that skin-colour tones facilitate detection, but only when this is tied to the general shape of a head. Face detection is impaired, for example, when faces are rendered entirely in greyscale or unnatural colours, or when skin-colour tones are preserved in only part of a face (Bindemann & Burton, 2009). Detection performance declines also when the general shape of a face is disrupted by image scrambling (Hershler & Hochstein, 2005). In contrast, face detection appears to be unaffected by some dramatic transformations, such as the removal of the internal facial features (i.e. the eyes, nose, and mouth), provided that general face-shape and colour information is retained (Hershler & Hochstein, 2005).

Viewed together, these studies suggest that face detection might be underpinned by skin-coloured, face-shaped templates. Beyond these findings, however, the nature of such a template remains largely unexplored. One aspect, for example, that has been preserved in all previous studies in this field is the height-to-width ratio of faces. Considering the impoverished nature of facial stimuli that allow detection to proceed unhindered (e.g. Bindemann & Burton, 2009; Hershler & Hochstein, 2005), such natural aspect ratios might be particularly important for

detection. However, while this idea seems plausible, an interesting discrepancy exists that might also undermine this notion. In tasks that require the *identification* of faces, substantial geometric distortions, which dramatically disrupt the typical height-to-width aspect ratios of faces, do not appear to affect performance. For example, even when faces are stretched vertically to 150% (Bindemann, Burton, Leuthold, & Schweinberger, 2008) or 200% (Hole, George, Eaves, & Rasek, 2002) of their actual size, while the original horizontal dimensions are maintained, the speed and accuracy of recognition is unaffected. This suggests also that face perception can be remarkably insensitive to manipulations that grossly distort stimulus shape.

This chapter, therefore seeks to explore how face detection is affected by such geometric distortions, to further investigate the nature of the template that might be used for this process. For this purpose, observers were asked to locate faces in images of natural scenes in a paradigm that is adopted from previous studies (Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Burton & Bindemann, 2009). In contrast to these studies, faces were either presented with their original aspect ratios intact or these ratios were manipulated. The aim here was to examine whether this would affect the efficacy with which faces can be detected, by recording observers' eye movements and response times to faces. If so, this would suggest that these aspect ratios are an important dimension of a face detection template. In a series of three experiments, Experiment 6 explored the detection of faces in their natural height-to-width ratio against vertically stretched faces. Experiment 7 further investigated the effect of stretching by equating the surface area of unstretched and vertically stretched faces. Experiment 8 then compared vertically and horizontally stretched faces.

Experiment 6

Experiment 6 examined how vertical stimulus distortions affect face detection. In this experiment, observers searched natural visual scenes for frontal views of faces, which were either presented in their original aspect ratio or were stretched vertically to increase the height-to-width ratio. Two different stretch conditions were used. In these, either the original height of the face stimuli was preserved but the width was compressed by half, or the original face width was preserved but the height was increased to double. These two conditions therefore provide identical height to width ratios (of 2:1), but one is comparable to the original face stimuli by retaining their height, whereas the other retains their width. If detection operates on a face-template that is sensitive to the height-to-width ratio of faces, then such geometric distortions should impair detection. As a result, observers should be slower to fixate these stretched faces in visual scenes and to make appropriate detection responses.

Method

Participants

Twenty-seven undergraduate students (8 male, 19 female) from the University of Kent, with mean age of 19.7 years ($SD = 2.2$), participated in this experiment for course credit. All reported normal or corrected-to-normal vision.

Stimuli

The stimuli were adopted from previous detection studies (Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Burton & Bindemann, 2009) and consisted of 24-bit RGB photographs of 120 indoor scenes, which were taken inside houses,

apartments and office buildings. These scene images measured 1000 (H) x 750 (W) pixels at a resolution of 72 pixels/inch (subtending a visual angle of $30.5^\circ \times 23.8^\circ$ at a viewing distance of 60 cm). For each scene, four versions were prepared which were identical in all aspects, except for the following differences. Three of these versions contained a photograph of a frontal face. The faces shown in these scenes were of twenty unfamiliar models (ten male, ten female) of white Caucasian origin. To ensure that the face locations were unpredictable throughout the experiment, the scenes were divided into an invisible 3 x 2 grid of six equally-sized rectangles. Across the stimulus set, the faces were equally likely to appear in any of these regions.

Apart from these commonalities, the three versions of these face-present scenes differed in terms of the aspect ratio of the faces. In the *original* face condition, the height-to-width ratios of all faces were preserved. However, the size of the faces was varied across scenes, ranging from 36 (H) x 27 (W) pixels ($1.2^\circ \times 0.9^\circ$ of VA) for the smallest face photograph to 139 x 115 pixels ($4.7^\circ \times 3.9^\circ$) for the largest face image (mean face image dimensions, 58.7 x 47.2 pixels ($2.0^\circ \times 1.6^\circ$); SD, 19.4 x 16.2 pixels ($0.7^\circ \times 0.5^\circ$)). This was done to ensure that participants could not adopt a simple search strategy based on the size of the faces (see Bindemann & Burton, 2009). The height-to-width ratio of these faces was also calculated. Height was measured as the maximum vertical distance between the facial boundary of the chin and the top of the forehead, whereas width was defined as the maximum horizontal distance between the left and right facial boundary by the ears. Across the stimulus set, the height-to-width ratio ranged from 1.08 to 1.75, with a mean of 1.44 (SD = 0.11). This is consistent with the average height-to-width ratio of this ethnic group (Farkas et al., 2005).

In the other two versions of the face-present scenes, these faces were either stretched vertically to twice the original height (i.e. to be 200%), while the horizontal dimensions were preserved, in the *vertically stretched* condition, or were compressed horizontally by half (i.e. to 50%) while the vertical dimensions were preserved, in the *horizontally compressed* condition. These two conditions therefore provide equivalent height-to-width ratios, but either only match the height or width of the original face stimuli. These manipulations were applied to each of the 120 scenes, resulting in a total of 360 face-present displays. In addition, a fourth version of each scene image was created in which the faces were absent, yielding 120 face-absent scenes. Example stimuli can be seen in Figure 3.1.



Figure 3.1 Example stimuli for Experiment 6, depicting a scene without face (top left), and faces in the original (top right), horizontally compressed (bottom left), and vertically stretched condition (bottom right).

Procedure

In the experiment, participants' eye movements were tracked using an Eyelink II head-mounted eye-tracking system running at 500 Hz sampling rate and SR-Research ExperimentBuilder software. Viewing was binocular but only the participants' dominant eye was tracked. To calibrate the eye-tracker, the standard 9-point Eyelink procedure was used. Thus, participants fixated a series of nine targets on the display monitor. Calibration was then validated against a second presentation of these targets. If the latter indicated poor measurement accuracy (i.e. a mean

deviation of more than 1° of participants' estimated eye position from the target), calibration was repeated.

In the experiment, a trial began with an initial drift correction for which participants were required to focus on a central target. A scene stimulus was then shown until a response was registered. Participants were asked to decide whether a face was present or absent in the scene by pressing one of two possible buttons on a standard computer keyboard. Participants were informed in advance that the faces could appear distorted in these scenes. Regardless of this, participants were requested to respond as quickly and as accurately as possible to the faces.

A total of 360 trials was shown to each participant, which consisted of 240 face-absent trials and 120 face-present trials. For face-present trials, 40 scene stimuli were shown in each of the experimental conditions (original, vertically stretched, horizontally compressed). The scene stimuli were rotated around these conditions across participants, so that each scene was shown only once to an observer in any of the face-present conditions. However, the presentation of the scenes was counterbalanced across participants, so that each scene was equally likely to appear in any of the conditions over the course of the experiment. All trials were presented in a randomly intermixed order.

Results

To assess detection performance, observers' accuracy (%) and response times (median correct RTs) were analysed first. This data is provided in Figure 3.2 and shows that detection accuracy was comparable in the original and the vertically stretched condition but was reduced for horizontally compressed faces. These

observations were confirmed by a one-factor within-subject ANOVA which showed a main effect of face type, $F(2,52) = 100.31$, $p < 0.001$, $\eta_p^2 = 0.79$. Post-hoc comparisons using Tukey HSD test showed that accuracy was reduced for horizontally compressed faces compared to their original and vertically stretched counterparts, both $qs \geq 16.60$, $ps < 0.001$, $ds \geq 4.84$. In contrast, performance for original and vertically stretched faces did not differ, $q = 1.40$, $d = 0.51$.

Observers' response times revealed a similar pattern. A one-factor within-subject ANOVA also revealed a main effect of face type, $F(2,52) = 116.59$, $p < 0.001$, $\eta_p^2 = 0.82$. Tukey HSD test showed that original and vertically stretched faces were detected faster than horizontally compressed faces, both $qs \geq 16.80$, $ps < 0.001$, $ds \geq 3.40$. In addition, response times were faster to vertically stretched than original faces, but this differences was not reliable, $q = 3.35$, $d = 1.32$.

In addition, the median time that was required to first fixate the faces in the visual scenes was also analysed. These *search times* were calculated for correct trials only and provide a more direct index of the search effort that is required to detect a face than button presses (i.e. response times). These eye movements were pre-processed by integrating very short fixations (< 80 ms) with the immediately preceding or following fixation if it lay within one degree of visual angle. The rationale for this was that such short fixations typically result from false saccade planning (see Rayner & Pollatsek, 1989).

As expected, search times were considerably faster than observers' button presses but reveal a similar pattern, whereby face detection appeared to be impaired in the horizontally compressed condition (see Figure 3.2). Accordingly, a one-factor

within-subject ANOVA of this data showed a main effect of face type, $F(2,52) = 50.44$, $p < 0.001$, $\eta_p^2 = 0.66$, due to slower response to horizontally compressed faces than their original and vertically stretched counterparts, both $qs \geq 11.86$, $ps < 0.001$, $ds \geq 2.22$ (Tukey HSD). In contrast, the search times for the original and vertically stretched faces did not differ, $q = 0.84$, $d = 0.30$.

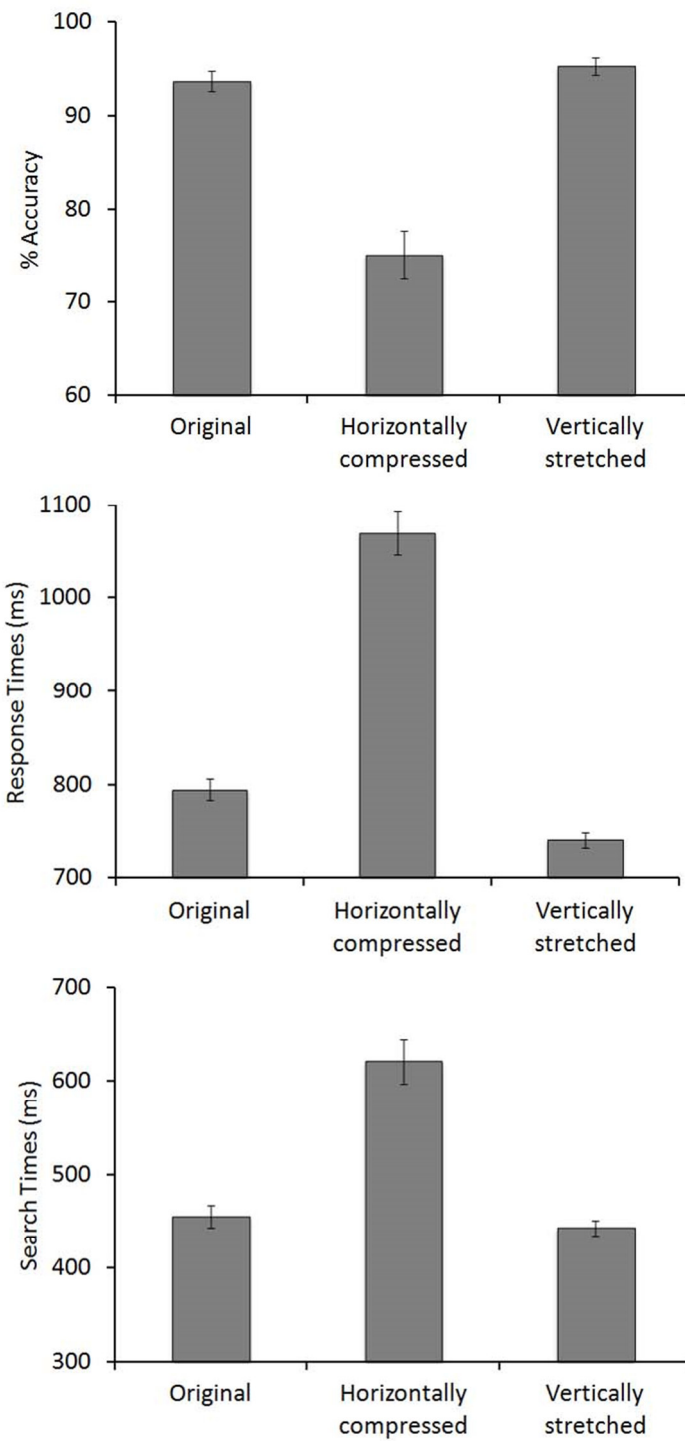


Figure 3.2 Detection accuracy (%), response times (ms), and search times (ms) for the face-present conditions in Experiment 6. Vertical bars represent the standard error of the means. Face-absent trials: accuracy = 99.0% (SE = 0.1), response times = 1813 ms (SE = 124).

Discussion

This experiment examined whether face detection is affected by the vertical distortion of faces. For this purpose, the detection speed and accuracy of unstretched faces, which were presented in their original dimensions, was compared with faces that were stretched vertically or compressed horizontally. Stretching impaired both the speed and accuracy of face detection. However, this effect was obtained only for faces that were “stretched” by compressing their width. In contrast, when faces were stretched to twice their original height, they were detected as well as their unstretched counterparts.

These results therefore appear to be inconclusive regarding the effect of stretching on face detection. However, a simple explanation might exist for the discrepancy between the horizontally compressed and the vertically stretched condition. These conditions were designed to be comparable to the original stimuli by retaining either the height (in the horizontally compressed condition) or width (in the vertically stretched condition) of these faces. As a result of this manipulation, however, the faces in the different detection conditions differ in terms of their surface area. In the horizontally compressed condition, for example, this area is reduced to half of the original face stimuli, with a corresponding increase in the vertically stretched condition. Surface area is known to affect face detection, whereby smaller faces are more difficult to detect than large faces (Bindemann & Burton, 2009). This raises the possibility that the effect of face stretching was masked in Experiment 6 by the differences in surface area between conditions. It is conceivable, for example, that the detection of vertically stretched faces was also impaired compared to the

unstretched originals, but this effect was offset by the increase in surface area in the former condition. This possibility is explored in Experiment 7.

Experiment 7

In Experiment 6, face detection was impaired for horizontally compressed faces, but not for faces that were stretched vertically. These conditions were matched in terms of their height-to-width ratio but differed in the surface area of the face stimuli. This raises the possibility that the effects of face stretching were offset by differences in area size. To dissociate the effects of surface area and stretching, face detection was assessed with four new conditions in Experiment 7. These comprised two conditions in which the original height-to-weight ratios of faces were retained. However, in one of these conditions the faces were presented at the same size as in Experiment 6, while, in the other, the size of the faces was increased to double their surface area. The faces were compared with two stretched conditions. Both of these provided altered height-to-width ratios by stretching faces vertically by 100% relative to the horizontal dimension. However, in one condition, the overall size of the stretched faces was adjusted so that the surface area was equated with the original face stimuli, whereas, in the other, surface area was also doubled. In line with previous findings, a detection advantage was expected for the large face conditions (see Bindemann & Burton, 2009). In addition, if stretching exerts an effect that operates independent of size, then face detection should be impaired in the stretched face conditions.

Method

Participants

Twenty-four undergraduate students (1 male, 23 female) from the University of Kent, with a mean age of 20.1 years ($SD = 3.8$), participated for course credits. None of them had participated in Experiment 6 and all reported normal or corrected-to-normal vision.

Stimuli and procedure

The stimuli were identical to Experiment 6, except for the following changes. In this experiment, four face-present scenes were included. These consisted of the original face stimuli (in the *original* condition) and a corresponding set of scenes, in which the height-to-weight aspect ratio was retained but the size of the faces was adjusted to double the surface area (in the *original large* condition). In addition, two stretched versions were created, in which the height-width ratio was increased by stretching faces vertically by 100% relative to the horizontal dimension. However, in one of these conditions, the face dimensions were adjusted further so that the surface area matched that of the original faces (in the *stretched* condition). In the other condition, stimulus size was increased so that surface area was at twice its original size (in the *stretched large* condition). Applying these manipulations to the 120 original face-present scenes resulted in a total of 480 experimental displays. Example stimuli are shown in Figure 3.3



Figure 3.3 Example stimuli for Experiment 7, depicting faces in the original (top left), original large (top right), stretched (bottom left), and stretched large condition (bottom right).

As in Experiment 6, each participant was shown 360 trials in a randomly intermixed order, comprising 120 face-present and 240 face-absent scenes. The face-present trials consisted of 30 scenes in each of the four experimental conditions (original, original large, stretched, stretched large). As in Experiment 6, the stimuli were rotated around these conditions across observers, but each scene was equally likely to appear in each condition over the course of the experiment.

Results

The data was analysed as in Experiment 6 and is provided in Figure 3.4. Accuracy was generally higher in the unstretched than the stretched conditions, and also when the surface area was increased to twice the original size. A 2 (face type: original vs. stretched) x 2 (face area: original vs. large) ANOVA showed a main effect of face type, $F(1,23) = 30.64, p < 0.001, \eta_p^2 = 0.57$, a main effect of face area, $F(1,23) = 46.12, p < 0.001, \eta_p^2 = 0.67$, and an interaction between both factors, $F(1,23) = 8.51, p < 0.01, \eta_p^2 = 0.27$. Analysis of simple main effects revealed an effect of face type for targets with the original area, $F(1,23) = 39.91, p < 0.001, \eta_p^2 = 0.63$, but not for the two large-area conditions, $F(1,23) = 2.28, p = 0.14, \eta_p^2 = 0.09$. In addition, a simple main effect of face area was found for original, $F(1,23) = 9.85, p < 0.01, \eta_p^2 = 0.30$, and stretched faces, $F(1,23) = 41.80, p < 0.001, \eta_p^2 = 0.65$.

Response times were analysed next. An analogous 2 x 2 ANOVA of this data also showed a main effect of face type, $F(1,23) = 27.03, p < 0.001, \eta_p^2 = 0.54$, a main effect of face area, $F(1,23) = 128.90, p < 0.001, \eta_p^2 = 0.85$, and an interaction between factors, $F(1,23) = 5.85, p < 0.05, \eta_p^2 = 0.20$. Analysis of simple main effects showed an effect of face area for original, $F(1,23) = 33.65, p < 0.001, \eta_p^2 = 0.59$, and stretched

faces, $F(1,23) = 105.29$, $p < 0.001$, $\eta_p^2 = 0.82$. These were complemented by simple main effects of face type for faces in their original size, $F(1,23) = 24.57$, $p < 0.001$, $\eta_p^2 = 0.52$, and in a large size, $F(1,23) = 5.74$, $p < 0.05$, $\eta_p^2 = 0.20$.

The analysis of eye movements also showed a main effect of face type, $F(1,23) = 15.51$, $p < 0.001$, $\eta_p^2 = 0.40$, due to faster search times for unstretched faces, and a main effect of face area, $F(1,23) = 47.51$, $p < 0.01$, $\eta_p^2 = 0.67$, with faster search times for the larger faces. The interaction between factors was not significant, $F(1,23) = 0.17$, $p < 0.68$, $\eta_p^2 = 0.01$.

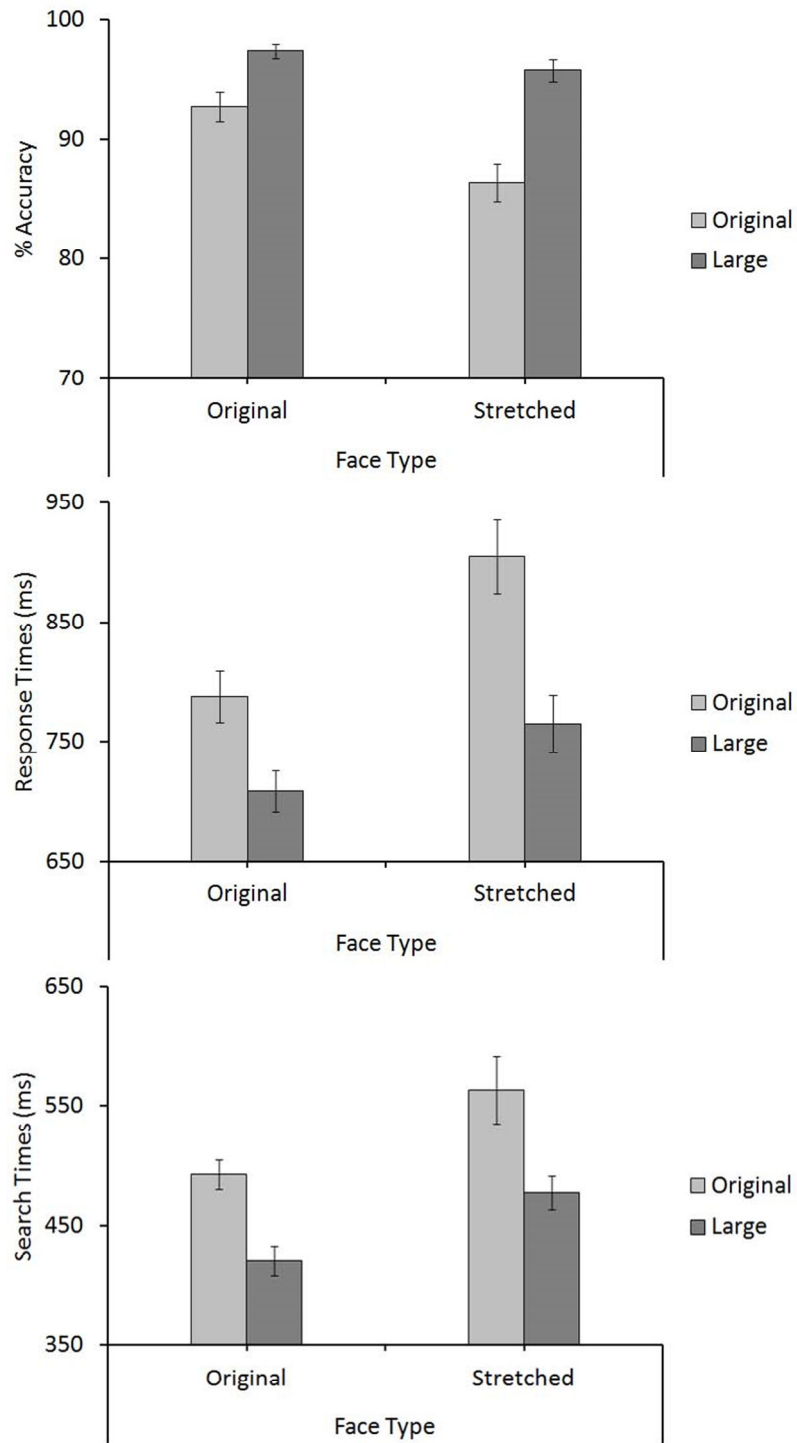


Figure 3.4 Detection accuracy (%), response times (ms), and search times (ms) for the face-present conditions in Experiment 7. Vertical bars represent the standard error of the means. Face-absent trials: accuracy = 99.0% (SE = 0.2), response times = 1666 ms (SE = 119).

Discussion

To provide a stronger test for the notion that face detection is affected by vertical distortions, the surface areas of unstretched and stretched faces were equated in Experiment 7. Moreover, to assess whether the effects of stretching and area are dissociable, two conditions were included, in which the original surface area of the face stimuli was either preserved or doubled. In line with previous work, a clear effect of face area was found, whereby both unstretched and stretched faces were detected faster in the large area conditions (see Bindemann & Burton, 2009). In addition, a separate effect of stretching was found, whereby faces were detected faster in their original height-to-width ratios than in the stretched conditions. This was evident in response times and eye movements, which indicates that this effect arises during the search for faces.

These findings help to clarify the results of Experiment 6. In that experiment, the stretched faces were equated to their original counterparts either in terms of their height or width. However, this manipulation also resulted in unequal surface areas for the faces across all conditions. As a consequence, it was impossible to separate the effect of face area from stretching. In contrast, Experiment 7 shows clearly that stretching impairs detection performance when the surface area of faces is controlled across conditions. In contrast to face recognition, which appears to be unaffected by the same geometric distortions (Bindemann, Burton, Leuthold, & Schweinberger, 2008; Hole, George, Eaves, & Rasek, 2002), these results suggest that detection relies on a template that incorporates the typical height-to-width aspect ratios of faces. So far, however, the current experiments have explored this notion only with vertically

stretched faces. In a final experiment, vertically and horizontally stretched faces are compared.

Experiment 8

In contrast to the preceding experiments, which compared faces in their original aspect ratios with vertical stretches, the current experiment included faces that were also stretched horizontally by 100%, to twice of the original face width. Face recognition appears to be unaffected by both types of stretches (Bindemann, Burton, Leuthold, & Schweinberger, 2008; Hole, George, Eaves, & Rasek, 2002). In turn, it is important to assess whether detection is only impaired by vertical or also by horizontal distortions of the typical height-to-width aspect ratios of faces.

Method

Participants

Thirty-two undergraduate students (3 male, 29 female) from the University of Kent, with a mean age of 19.3 years ($SD = 1.0$), participated for course credits. None of these students had participated in the preceding experiments. All reported normal or corrected-to-normal vision.

Stimuli and procedure

The stimuli and procedure were identical to Experiment 7, except for the following changes. In addition to the 120 original face-present scenes, in which faces were presented in their natural height-to-width ratio, two more versions were created of each scene. One of these versions consisted of vertically-stretched faces from

Experiment 7, which matched the surface area of the original faces. The other version consisted of horizontally-stretched faces. These faces were prepared in the same manner as their vertically-stretched counterparts, except that the opposite height-to-width ratio was used. This resulted in a total of 360 displays, comprising 120 scenes for each of the face-present conditions (original, vertically stretched, horizontally stretched). Example stimuli are shown in Figure 3.5.

In the experiment, each observer was shown 240 face-absent and 120 face-present displays (40 displays for each of the original, horizontal stretched and vertical stretched faces) in a randomly-intermixed order. As in previous experiments, the face stimuli were rotated around the three face-present conditions across observers, so that each face-present scenes was only encountered once, but all scenes were equally likely to appear in each of the face conditions over the course of the experiment.



Figure 3.5 Example stimuli for Experiment 8, depicting faces in the original (top), horizontally stretched (middle), and vertically stretched condition (bottom).

Results

The data from one participant, whose search times were more than five standard deviations from the group mean, was excluded from all analysis. For the remaining 31 observers, accuracy, reaction times and search times are shown in Figure 3.6 A one-factor within-subject ANOVA showed a main effect of face type, $F(2,60) = 9.85, p < 0.05, \eta_p^2 = 0.25$. Tukey HSD test shows that this reflects reduced detection accuracy for vertically and horizontally stretched faces compared to their original counterparts, both $qs = 5.44, ps < 0.001, ds \geq 1.12$, while the two stretched conditions did not differ from each other, $q = 0.00, d = 0.00$.

A similar effect of face type was also found for response times, $F(2,60) = 26.63, p < 0.001, \eta_p^2 = 0.47$, and search times, $F(2,60) = 16.01, p < 0.001, \eta_p^2 = 0.35$. For both measures, Tukey HSD showed that the original faces were detected faster than their vertically and horizontally stretched counterparts, all $qs \geq 5.96, ps < 0.001, ds \geq 1.32$. In both response and search times, the two stretched conditions did not differ from each other, both $qs \leq 1.65, ds \leq 0.31$.

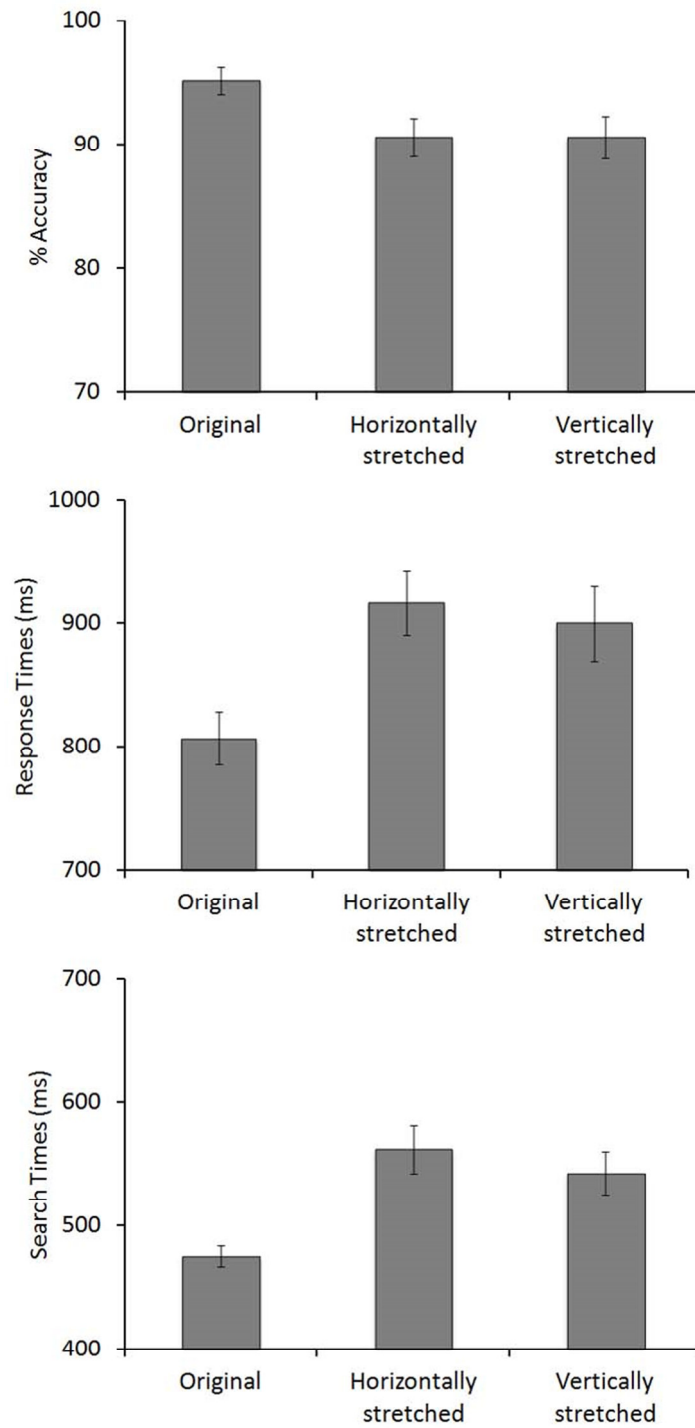


Figure 3.6 Detection accuracy (%), response times (ms) and search times (ms) for the face-present conditions in Experiment 8. Vertical bars represent the standard error of the means. Face-absent trials: accuracy = 99.0% (SE = 0.2), response times = 2007 ms (SE = 137).

Discussion

The results of this experiment confirm that face detection is affected by vertical distortions and extend this finding to horizontally stretched faces. As in Experiment 7, this effect was found despite the fact that these stretched faces matched the surface area of their unstretched counterparts. This finding then suggests that face detection relies on a template that utilizes typical height-to-width aspect ratios of faces. These findings are discussed in the General Discussion.

General Discussion

This study examined whether geometric distortions, by stretching faces to manipulate their natural height-to-width aspect ratio, impairs person detection. The impact of stretching on detection performance was not obvious when faces were equated to their original, unstretched counterparts in terms of their height or width dimension (Experiment 6). However, a clear effect of stretching was obtained when the original and distorted faces were matched for their surface area (Experiment 7), and this was found for both vertically and horizontally stretched faces (Experiment 8). This effect was evident in the accuracy and speed of observers' detection responses and also in the initial eye movements to faces, which indicates that it arises during the *search* for faces in natural scenes. Moreover, this effect was found despite the fact that observers were informed of the stretched face conditions prior to the experiment. Taken together, these results suggest that the effect of stretching on face detection is remarkably robust.

The question arises of which aspect of faces changes to impair detection when these stimuli are stretched vertically or horizontally. One possibility is that this manipulation distorts internal facial features, or the distances between these, which then impairs detection. However, scrambling or removal of internal features only appears to impair face detection when the outline of faces is also disrupted and these are presented in unnaturalistic scenes (see Garrido, Duchaine, & Nakayama, 2008; Lewis & Edmonds, 2003, 2005). By contrast, detection proceeds unimpaired when internal features are scrambled or removed, provided that a general face outline is preserved.

In conjunction with the findings of the experiments presented here, this suggests that the height-to-width aspect ratio of faces is a specific component of the cognitive template that is utilized for detection. The findings from Chapter 2 already suggest that this template might rely on a “quick and dirty” processing strategy that utilizes some salient but simple visual cues to locate likely face candidates. These simple cues containing only colour and face shape has been shown to help rapid face search but higher detail structure of the eyes, the nose and mouth delay detection. This addresses the importance of skin colour face shape template for rapid detection. In consistent with previous detection studies, it has been shown, for example, that detection proceeds unhindered when internal (i.e. eyes, nose and mouth) or external facial features (e.g. face outline, hairstyle) are removed, as long as an oval face-shaped template is preserved (Hershler & Hochstein, 2005). Face detection is also facilitated by skin-colour tones but only when these are tied to the shape of a face (Bindemann & Burton, 2009). In contrast, detection performance is impaired when overall face-shape is destroyed by image scrambling (Hershler & Hochstein, 2005) or

bit-part deletion (Burton & Bindemann, 2009). Taken together, these results indicate that face detection might be driven by a simple skin-coloured face-shape template. The experiments in this chapter add to these findings by suggesting that this template utilizes the natural height-to-width ratio of faces to aid detection.

To explore the role of such aspect ratios for face detection, the current study stretched faces vertically or horizontally to 200% of their original size, while maintaining the size of the orthogonal dimension. While this is a dramatic transformation, the question arises of whether the cognitive detection template is sensitive to smaller distortions that reflect natural between-subject variation of facial height-to-width ratios. To begin to explore this *a posteriori*, the response times to the original faces were calculated across all three experiments as a function of their height-to-width ratio. While these ratios ranged from 1.08 to 1.75, only very few faces had such extreme ratios. The stimuli were therefore divided into larger non-overlapping face categories with height-to-width ratios that were close to 1.2, 1.4, 1.6 and 1.8. A one-factor ANOVA of this data, which is illustrated in Figure 3.7, showed an effect of ratio, $F(3,269) = 14.48$, $p < 0.01$ ¹, which reflects slower responses to faces in the 1.2 and 1.8 categories than for the two intermediate face ratios (Tukey HSD, all $ps < 0.01$). A similar pattern was obtained for search times, $F(3,241) = 3.45$, $p < 0.05$, which were slower for the 1.8 than the 1.6 and 1.4 categories (both $ps < 0.05$), while faces with a 1.2 ratio did not differ from any of the categories. Overall, these data therefore suggest that face detection is best with height-to-width ratios in

¹ Some participants failed to record a single correct response in some of the height-to-width categories. Because of these missing data points, ANOVA was computed on a between-subjects basis.

the range of 1.4 to 1.6. This conclusion is drawn tentatively, as these ratios were not manipulated systematically across our scenes.

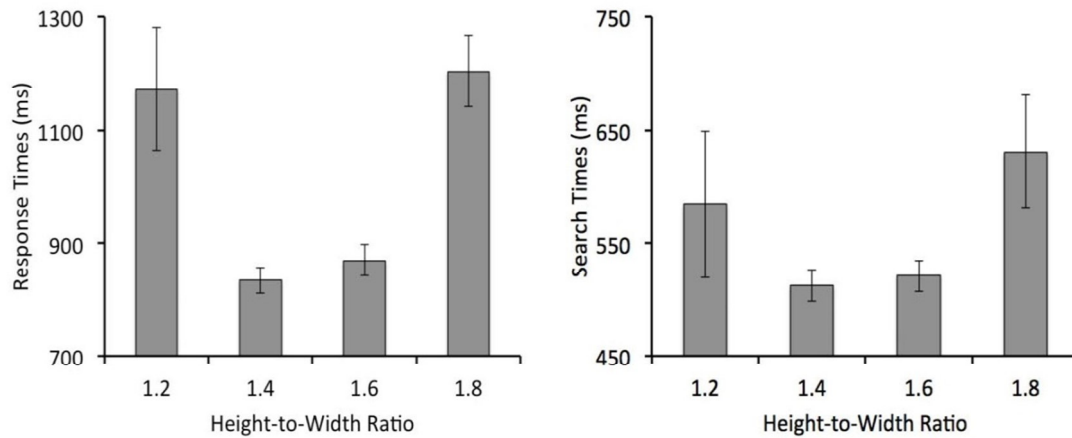


Figure 3.7 Response times (ms) and search times (ms) for the original face stimuli in Experiments 6 to 8, grouped by height-to-width ratio. Vertical bars represent the standard error of the means.

The effect of geometric distortions on face detection is interesting considering that observers appear insensitive to subtle differences in the height-to-width ratio of individual face *identities* (Sandford & Burton, 2014), and as person *recognition* is also unaffected by the drastic manipulations that impaired the detection of faces in the current experiments (see, e.g. Bindemann, Burton, Leuthold, & Schweinberger, 2008; Hole, George, Eaves, & Rasek, 2002). This differential sensitivity to geometric distortions converges with other recent findings to indicate that detection differs from other tasks with faces (Bindemann & Lewis, 2013). In this respect, it is interesting to note that face detection might also differ from the perception of non-face stimuli, such as natural and urban scenes, which also appear to be insensitive to substantial linear distortions (e.g. up to 52%, see Kingdom, Field, & Olmos, 2007; see also Cutting, 1987).

Chapter 4:

Summary, Conclusions and Future Research

4.1 Summary and Conclusions

This thesis investigated how human observers detect faces in complex natural scenes. The process of face detection appears to be distinct from other tasks with faces (Bindemann & Lewis, 2013). However, in contrast to face tasks such as identification (see, e.g. Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990; Burton, Jenkins, Hancock, & White, 2005), matching (Burton, White, & McNeill, 2010; Clutterbuck & Johnston, 2002; Johnston & Bindemann, 2013), and emotion recognition (e.g. Calder, Burton, Miller, Young, & Akamatsu, 2001; Calder & Young, 2005), detection has been studied comparatively little. The available evidence demonstrates that detection is rapid, so that eye movements to faces are initiated in just 100 ms (Crouzet, Kirchner, & Thorpe, 2010; Fletcher-Watson, Findlay, Leekam, & Benson, 2008), and automatic (Lewis & Edmonds, 2003, 2005; Crouzet et al., 2010). This suggests that face detection might be driven by a ‘quick and dirty’ processing strategy that relies on simple visual cues (Crouzet & Thorpe, 2011). One possibility is that such a strategy is based on a skin-coloured face-shape template (Bindemann & Burton, 2009; Hershler & Hochstein, 2005).

This idea derives from the observation that detection is impaired when faces are rendered in greyscale, appear in unnatural colours through hue-reversal, or when colour information is preserved in only part of a face (Bindemann & Burton, 2009). Similarly, detection performance declines when faces are presented in their natural colours but shape information is disrupted through scrambling (Hershler & Hochstein, 2005), inversion (Garrido, Duchaine, & Nakayama, 2008), or part deletion (Burton & Bindemann, 2009). This indicates that shape or colour alone cannot account for optimal performance in face detection, but work in conjunction.

However, other facial information appears to be important for detection too, as faces are still detected when shape and colour information is disrupted. Under these conditions, detection appears to be supported by features, such as the eyes (Burton & Bindemann, 2009). Indeed, a simple configuration of four dots, to represent two eyes above a central nose and mouth, is enough to guide attention (Johnson, Dziurawiec, Ellis, & Morton, 1991; Macchi, Simion, & Umiltà, 2001) and initiate responses to face-like regions (Awasthi, Friedman, & Williams, 2011b, Nestor, Vettel, & Tarr, 2013; Simion, Farroni, Cassia, Turati, & Barba, 2002; Simion, Macchi, Turati, & Valenza, 2003). Similarly, a simple black-white contrast of Mooney faces, which contain featural patterns, can be accounted into a face (Andrew & Schluppeck, 2004; George, Jemel, Fiori, Chaby, & Renault, 2005). By contrast, disruption of such a pattern through inversion (Garrido, Duchaine, & Nakayama, 2008) or feature scrambling delays performance (Hershler & Hochstein, 2005). Thus, internal features also help to support detection. However, the question arises of how salient visual cues, such as general shape and colour information, and more detailed cues, such as features, are prioritized for detection.

This thesis examined this question over a series of eight experiments, by applying several new manipulations. Chapter 2 began by investigating how shape and features are integrated into a face-template during detection by isolating information from different spatial frequencies (SF). Specifically, observers were asked to find faces in scenes that were presented in their original, intact spatial frequency content or in which only low (LSF), mid (MSF) or high (HSF) spatial frequencies were preserved. These conditions either presented large-scale luminance variation (in LSF) and should therefore preserve salient visual cues only, such as colour and shape, or

small-scale luminance variation (in HSF), which preserves finer visual detail, such as the eyes, nose and mouth. MSF provided an intermediate level of detail, between these two spatial bandwidths.

By comparing performance for these SF conditions, the experiments in Chapter 2 sought to determine which information affects face detection. Thus, if detection is driven by a 'quick and dirty' processing strategy that utilizes only simple visual cues such as a skin-coloured face-shaped template, then detection performance should have been best for LSF. In contrast, if fine visual details, such as that contained in internal facial features, are important for detection, then HSF should have been the most useful condition for this purpose. Finally, if detection requires the combination of both shape and feature information, then MSF, which provides an intermediate level of detail (i.e. gross detail of shape and features), should have yielded the best results.

In Experiment 1, SF was manipulated by filtering the entire scene area. Accuracy, response times, and eye-movements were recorded to assess face detection. Under these conditions, detection accuracy was best for faces in the original and MSF scenes, and lowest for LSF and HSF scenes. In contrast, faces were detected fastest in the original condition, but both MSF and LSF faces were also detected faster than in the HSF condition. The eye movement data confirmed these response-time findings by showing the same pattern in search times. The accuracy, response times and search times therefore consistently showed best detection for the original scenes, followed by MSF scenes, and worst detection for HSF. By contrast, these measures provided conflicting results for the LSF condition, as these faces were detected with low accuracy but, when detected, they were responded to quickly.

However, because the entire scene stimuli were filtered to produce the different SF conditions in Experiment 1, it was possible that the pattern of results might not reflect the effect of different SF on face detection but might reflect scene processing instead. To explore this issue, SF was only applied to the face photographs that were embedded within these scenes in Experiment 2, whereas the scene background was now left intact. And in Experiment 3, SF was only applied to the face image within these photographs. In both experiments, a similar pattern to Experiment 1 was found. Thus, the original faces were detected best, both in terms of accuracy, response and search times, and performance was intermediate for MSF and worst for HSF faces. As in Experiment 1, the LSF faces were also detected as quickly as MSF faces in Experiment 2 and 3. In contrast to Experiment 1, however, LSF faces were also detected more accurately than the HSF condition in these experiments. Thus, detection accuracy for LSF faces was improved by manipulating the SF content of the faces in the scenes only. Taken together, these findings further support the notion that an intermediate level of detail, such as MSF, is best for face detection. However, LSF faces are detected as quickly as MSF, which suggest that this SF band in particular supports the *fast* detection of faces. Occasionally, however, LSF information also can be more limiting than MSF and lead observers to miss faces entirely in complex natural scenes.

The original scene stimuli in Experiments 1 to 3 were always shown in their natural colours. However, during filtering this colour information is preserved only in LSF, whereas MSF and HSF faces are essentially rendered in greyscale. As colour faces are easier to detect than greyscale faces (see Bindemann & Burton, 2009; Lewis & Edmonds, 2003, 2005), this raised the possibility that the performance in the SF

conditions was determined by different cues. For LSF faces, this might have reflect the available colour information in Experiments 1 to 3 rather than SF content. The detection of MSF and HSF faces, on the other hand, might have been supported by the SF content of these stimuli rather than colour cues. To explore this possibility, Experiment 4 assessed face detection by rendering all stimuli in greyscale.

Despite the removal of colour information, Experiment 4 replicated the key findings of the preceding experiments, by showing that detection was fastest and most accurate for original and MSF faces and worst for faces in HSF. Most importantly, detection of LSF faces was still as fast as MSF, both in terms of response and search times. This finding converges with the preceding experiments and demonstrates that the speed advantage of LSF faces is not determined simply by colour content, but must be related more directly to the SF.

The final experiment of this chapter then investigated whether the current findings could reflect an artefact that arises from the manipulations that were employed in Experiments 2 to 4. In these experiments, the SF content of the face regions was selectively filtered while the scene background remained intact. This raised the possibility that these scene regions somehow stood out to observers and attracted their attention, rather than the face content of these regions *per se*. To investigate this possibility, which receives support from the finding that small blurred regions within images can attract observer's fixations (Smith & Tadmor, 2013), Experiment 5 included additional face-absent conditions in which patches of LSF and HSF content were embedded within scenes. These SF patches matched the location and size of the faces in the face-present counterparts.

Consistent with all of the preceding experiments, the original faces were detected better than LSF faces whereas performance was worst for HSF faces. Crucially, however, the face-absent scenes with the LSF patches were not classified more accurately or faster, and the patches were not fixated quicker, than those in HSF scenes. Moreover, the percentage of trials in which these regions were fixated was low compared to face-present scenes. These findings therefore suggest that the results of Experiments 2 to 4 are not simply an artefact of the experimental manipulations but reflect the role of specific spatial frequency for face detection. Overall, the experiments in Chapter 2 therefore provide consistent evidence that LSF and MSF information support the rapid detection of faces.

While the experiments in Chapter 2 suggest that LSF information is particularly useful for the fast detection of faces, the question arises of which information is preserved in LSF faces that could drive such effects. It is already known that colour information facilitates detection, so this could be one of the visual characteristics that drives such effects (Bindemann & Burton, 2009). However, the detection of LSF faces also remained fast when colour information was removed in Experiment 4. The detection of LSF faces must therefore be facilitated by further information. Chapter 3 examined whether this information might reflect the basic dimensions of a face-shape template, such as its general height-to-width ratio.

To investigate this question, geometric distortions were applied to the faces in the scenes, by stretching these stimuli selectively in either a vertical or horizontal plane, while the orthogonal dimension remained intact. If such geometric distortions delay detection, then it would suggest that the normal height-to-width ratio of faces is an important aspect of the cognitive template for detection. In Experiment 6,

observers searched for faces that were either stretched vertically to twice the original height (i.e. to be 200%), while the horizontal dimensions were preserved, in the *vertically stretched* condition, or were compressed horizontally by half (i.e. to 50%) while the vertical dimensions were preserved, in the *horizontally compressed* condition. These stretching manipulations were derived from a previous study on face recognition (Hole, George, Eaves, & Rasek, 2002) and were compared with an original condition, in which faces were shown in their actual height-to-width ratios. In comparison to the original condition, the vertically stretched and horizontally compressed conditions provide identical height-to-width ratios (of 2:1), but one was comparable to the original face stimuli by retaining their height whereas the other retained their width.

In Experiment 6, stretching impaired the speed and accuracy of face detection, but this effect was observed only with faces that were manipulated by compressing their width. By contrast, vertically stretched faces were detected as well as their unstretched counterparts. These results were therefore inconclusive regarding the effect of stretching on face detection. However, while these conditions were designed to be comparable to the original stimuli by retaining either their height (in the horizontally compressed condition) or width (in the vertically stretched condition), the faces differed in terms of their surface area across conditions. As this is known to affect face detection, whereby smaller faces are more difficult to detect than large faces (Bindemann & Burton, 2009), further experiments were conducted. Experiment 7 altered height-to-width by stretching faces vertically by 100% relative to the horizontal dimension, but also controlled the surface area across the original and the stretched condition. Two different surface areas were applied. In one, the overall size

of the stretched faces was adjusted so that the surface area was equated with the original face stimuli. In the other, the surface area of faces was doubled, both in the original and the stretched condition.

In support of previous work, a clear effect of surface area was found, such that unstretched and stretched faces were detected faster in the larger area conditions (see Bindemann & Burton, 2009). This was accompanied by an effect of stretching, whereby faces were detected slower in the stretched conditions than when their original height-to-width ratios were preserved. Crucially, this effect of stretching held when surface area was controlled across conditions and, as indicated by response times and eye movements, arose during the search for faces. These results were confirmed by a final experiment that compared detection performance for original, vertically and horizontally stretched faces (Experiment 8). Together, these experiments clearly show that these geometric distortions disrupt face detection. In turn, these findings suggests that detection relies on the typical height-to-width aspects of faces.

These findings are interesting considering that face recognition appears to be completely unaffected by similar distortions (Bindemann, Burton, Leuthold, & Schweinberger, 2008; Hole et al., 2002) and converge with recent claims that detection is separable from other tasks of faces (Bindemann & Lewis, 2013). However, while the experiments in Chapter 3 reveal the importance of retaining the general height-to-width ratios of faces to optimize detection, the results also suggests that detection is sensitive to smaller distortions of these ratios that reflect between-subject variation in identity. For the face stimuli used in the current experiments, the ratios for the original face stimuli ranged from 1.08 to 1.75. Within this range,

detection performance was best for faces with a ratio of between 1.4 and 1.6, whereas more extreme ratios, outside of this range, appeared to delay observers' eye movements to faces and their responses. A potential limitation of this finding is that only very few face stimuli displayed more extreme height-to-width ratios. Thus, further work is needed to explore this particular issue more thoroughly.

A number of conclusions can be drawn from the experiments in Chapter 2 and 3. Firstly, the finding that LSF information supports the detection of faces in Chapter 2 is consistent with the notion that this process is driven by low-level information (Crouzet & Thorpe, 2011), such as a simple face-shape template. Previous research suggests that face detection remains fully effective when a simple oval face-shape is preserved, even when facial features, such as the eyes, nose and mouth are removed (Hershler & Hochstein, 2005). The current results converge with these findings by showing that highly blurred (LSF) stimuli, which only preserve the broadest detail of a face, are sufficient for *fast* detection. However, the experiments in Chapter 3 suggest that the original height-to-width ratios of faces might form part of such a detection template.

Secondly, while detection was very fast for LSF faces, these faces were missed more frequently than the MSF and original faces in Chapter 2. Thus, LSF information is not sufficient to always support *accurate* detection. Moreover, even when LSF is removed – for example, in the HSF condition – the majority of faces were still detected in the visual scenes. Similarly, while performance was impaired for stretched faces in Chapter 3, these were still detected on the majority of trials (i.e. > 75%) and, on average, in under a second. These findings indicate clearly that additional sources of information, other than a simple LSF face-shape, contribute to

detection. The visibility of the eye regions in a face might provide one source of such alternative information (see Burton & Bindemann, 2009).

Thirdly, these findings suggest that the *fast* detection of faces is supported by the ‘quick and dirty’ processing channel of the magnocellular pathway (subcortical route), which specifically supports the processing of LSF (Bullier, 2001; Livingstone & Hubel, 1988). This channel also appears to be suited best for processing faces in the visual periphery (Awasthi, Friedman, & Williams, 2011a, 2011b), which is important for detection. By contrast, the fine visual detail of HSF is likely to be processed via the fovea-sensitive parvocellular channel in ventral stream (Bullier, 2001; Livingstone & Hubel, 1988; Lynch, Silveira, Perry, & Merigan, 1992). This slower channel might help to maximise performance in the original and MSF conditions, or when LSF information is sub-optimal or unavailable, such as in HSF displays.

One way to integrate these observations could be a detection model in which the two channels - the magnocellular and the parvocellular brain pathways - process faces *in parallel* but at different speeds in a horse-race model. According to such a model, salient visual content, such as LSF information and basic height-to-width ratios, is normally processed faster and therefore provides a detection advantage in terms of speed. However, when such information is compromised – for example, when faces with highly unusual height-to-width ratios are encountered – other visual content, such as the finer visual detail of HSF, becomes the primary information source for detection. This model could explain why detection is fastest for LSF but also why faces can still be detected from the slower and less accurate HSF cues. The combination of different information sources, such as LSF and HSF information, within such a model could also explain why performance is maximised in the original

and MSF face conditions (which contain both LSF and HSF ranges or intermediate content). In further support of such a model, it has already shown that early face processing, which is reflected in the N170 event-related potential, is best when both LSF and HSF information are available (Halit, de Haan, Schyns, & Johnson, 2006).

The current experiments might also serve to rule out some alternative models of face detection. For example, it is possible that detection might reflect a sequential two-stage process, comprising the initial search for likely face candidates and a subsequent decision stage to confirm that a located stimulus is, in fact, a face. According to such an account, LSF might support the fast localization of likely face-shapes whereas HSF is then utilised for confirmation. However, several aspects of the current data argue against such a sequential two-stage model. In such a framework, one might expect that faces are occasionally fixated as likely face candidates, but not detected due to insufficient information to make a confirmatory response. This might be particularly the case for LSF faces, which can provide very limited visual information (see examples in Figures 2.6). If so, one would expect that these faces are frequently fixated but not detected. However, an analysis of such a measure failed to find an effect of condition in Experiments 2, 3 and 4 (see Figure 2.14).

Secondly, if such a sequential model can be applied, then one might expect that the time it took to first fixate a face in a scene provides a more direct measure of the search for likely face candidates, whereas the time taken to respond to a face is representative of a search and confirmatory decision process. By subtracting response from search times, a measure of the time it might have taken observers to reach such a confirmatory decision was derived. If the search for likely face candidates is a separable decision-making stage, then one might expect that these decision times

show an advantage for the finer visual detail carried by HSF information. Contrary to this notion, however, such an HSF advantage was consistently absent across experiments (see Figure 2.15). The current findings therefore do not appear to support a sequential two-stage model. Rather than a two-stage model, another explanation could incorporate the cumulative accrual of facial detail until a threshold to make a detection decision is reached would also be possible. Accordingly, only fragments of information, such as a pairs of the eyes within an oval shape as provided by MSF (Ullman et al, 2002) or general height-to-width shape information as provided by LSF, need to accrue to pass a threshold level of being a face as required by relevant subcortical neurons (see e.g. Johnston, 2005, Nguyen et al., 2013, 2014). In this framework, information might also be combined from multiple information sources and detectors.

In summary, this thesis explored the detection of faces in natural scenes by manipulating the SF content across five experiments in Chapter 2 and the height-to-width ratios of faces across three experiments in Chapter 3, and by recording observers' accuracy, response times, and eye movements. The main findings of these experiments are, firstly, that extremely impoverished LSF stimuli are sufficient for the fast detection of faces. However, occasionally this information is simply too limiting and can therefore lead observers to miss faces altogether. By contrast, face detection is slow from HSF content that preserves fine visual detail, but not impossible. Secondly, the natural height-to-width ratio of faces appears to be an important component of the cognitive template for face detection, as performance is impaired when these ratios are disrupted. Considering the extremely limited information that appears to be preserved in LSF, it is possible that height-to-width ratios are the key

information for detection that is preserved in these greatly impoverished stimuli. The experiments in this thesis stop short of examining this directly, by combining the SF manipulations and geometric distortions in the same experiment, but this is clearly an interesting question for further research.

In addition it would be interesting to further investigate the function of colour for face detection. While LSF can support the rapid detection of faces independently of colour, colour does help to maintain accuracy for LSF faces. This suggests that colour might be processed independently of face-shape module but both aspects can function together to improve performance. Future studies could investigate how colour and shape are integrated, by directly comparing detection performance for coloured LSF faces (intact colour and shape), greyscale LSF faces (intact shape only), and part-colour LSF faces (intact colour only). To this point, it is notable that colour perception is also mediated by the relatively slow parvocellular pathway (Shapley & Hawken, 2002). Consequently, it is possible that colour and shape might be processed in a similar parallel fashion as shape and features.

These findings could be integrated in a detection framework that combines salient visual cues, such as LSF and height-to-width ratios in a simple oval grey or skin-coloured template (see Bindemann & Burton, 2009), and finer visual detail from HSF, such as featural information from the eyes (see Burton & Bindemann, 2009), in a horse-race model. Such a model is proposed in Figure 4.1. According to this model, multiple sources of information might support detection in parallel depending on the respective availability of these different cues.

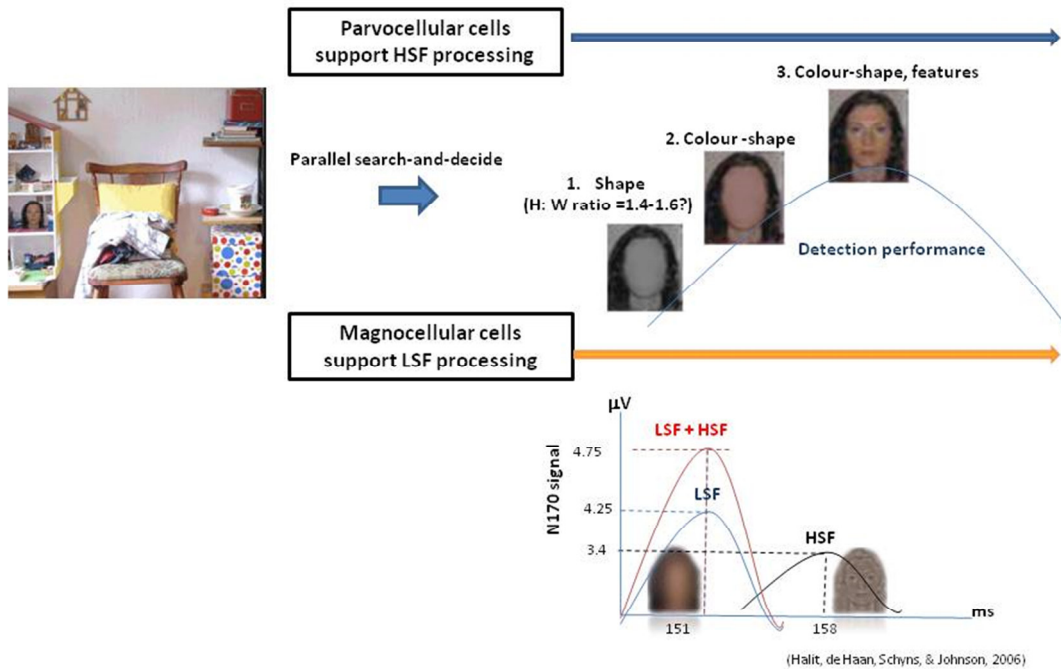


Figure 4.1 A horse-race model of face detection. Accordingly, face detection can be activated by both the magnocellular (achromatic) and parvocellular channel. The cognitive template for detection might include three main components, comprising face-shape with normal height-to-width ratio, colour information and facial features. Under normal conditions, detection is supported by both pathways and, either colour, -shape or feature information can drive detection (Hershler & Hochstein, 2005). Under poor acuity, detection is supported more by the magnocellular pathway, via LSF shape information. In contrast, when LSF information is suboptimal, detection utilises other sources of information, such as shading of the eyes, or simple feature patterns. Thus, the simplest cognitive template is the shape of the face, probably with a normal height-to-width ratio of 1.4-1.6. Detection performance can improve further through the addition of skin colour or feature cues.

4.2 Limitations and suggestions for future research

The current experiments raise several questions for further research. The first question concerns the sensitivity of the SF manipulation in Chapter 2. The cut-off values to filter stimuli in different SF bands (e.g. 5 cycles/face for LSF faces) were based on the mean dimensions of these faces across the set of 120 scenes. Within these scenes, however, faces varied in size. For example, whereas the mean size of faces was at 59 (H) x 47 (W) pixels (at a resolution of 66 ppi), the size of individual faces ranged from 36 x 27 to 139 x 115 pixels. This raises the question of whether the current effects hold when the filtering of SF information from faces is adjusted more finely to take account of individual stimulus size.

Another question concerns the conclusion that a sequential two-stage process, consisting of an initial search for likely-face candidates and a confirmatory stage that a looked-at stimulus is, in fact, a face, is unlikely to explain detection performance. While the current experiments measured observers' responses and eye movements to understand detection performance in depth, these conclusions were drawn with the caveat that there might be other paradigms that can examine such a two-stage account more directly. One possibility for such a paradigm could be a gaze-contingent method, in which the onscreen content is directly linked to location of observers' eye fixations (see Duchowski, Cournia, & Murphy, 2004). Such a paradigm could be designed to remove a face from a scene when observers' eyes move close to its location or, conversely, for the face to only appear when its location falls within observers' foveal vision. Thus, the face stimuli could be manipulated to be present onscreen either only during the search but not the decision process, or vice versa.

Another method that might be suitable for resolving whether separable search and decision processes exist could be eye fixation-related potentials (EFRP), which combine the measurement of eye movement and neural activity with EEG (Fischer, Graupner, Velichkovsky, & Pannasch, 2013). This cutting edge approach might allow the study of face detection by monitoring neural and oculomotor functions online. While this combination might allow to understand which neural processes trigger certain eye movement events (Fischer, Graupner, Velichkovsky, & Pannasch, 2013; Frey et al., 2013; Henderson, Luke, Schmidt, & Richards, 2013), this methodology has only received limited attention in the study of face detection. Existing studies have applied EEG measurement to study face processes such as searching and decision-making separately in saccadic choice (Kirchner & Thorpe, 2006) or judgement tasks (Halit et al., 2006). This research shows that EEG can provide insight into these processes in some paradigms when these are examined individually. By combining EEG with eye-tracking, it might also be possible to assess *both* processes more directly in future studies of face detection. It has already been demonstrated that such an EFRP approach can dissociate pre-defined events, such as the onset of stimuli, and un-defined events, such as the timing of decisions to a stimulus (Frey et al., 2013). In this context, undefined events such as decision-making could be defined by the time periods in which eye movements indicate that observers have stopped searching a visual display. EEG at this point in time might then indicate which SF information elicits the largest ERP components and is therefore most important for decision-making. It is therefore possible this approach can also provide further insight into the different stages and processes involved in face detection.

Another approach that could be used to study face detecting further is image averaging. In the study of face recognition, averages of specific facial identities have been created by morphing together multiple images of the same person's face (see Burton, Jenkins, Hancock, & White, 2005; Jenkins & Burton, 2008, 2011). These averages appear to be a good way of capturing the internal cognitive templates of individual identities (more refs – see Mike Burton's webpage for possibilities). However, whereas recognition requires the individuation of faces (i.e. this process must capture how the faces of different people differ), detection should build on the information that is *shared* across identities. Thus, individual averages might exist for all known facial identities, but only average for their detection. This would be an interesting question for further research that could be examined by exploring the detection of specific exemplars of faces (i.e. as in the current experiments), identity averages (as in Burton et al., 2005), and averages that combine identities.

Lastly, the current research might also provide insight into the development of face detection algorithms in computing science. The experiments in Chapter 2 suggest, for example, that template matching in computerized detection would be fastest if a simple LSF filter is used first to look for faces. Subsequently, mid-resolution filters could be used to identify facial fragments when low-level information is sub-optimal (for example, when faces in scenes are partially occluded) or for confirming the presence of a face. It would also be interesting to see how specific combinations of colour, shape, and height-to-width ratios improve face detection algorithms (see, e.g. Viola & Jones, 2004; Sinha, 2002).

References

- Aguado, L., Serrano-Pedraza, I., Rodríguez, S., & Román, F. J. (2010). Effects of spatial frequency content on classification of face gender and expression. *The Spanish Journal of Psychology, 13*, 525-537. doi:10.1017/S1138741600002225
- Andrews, T. J., & Schluppeck, D. (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. *Neuroimage, 21*, 91-98. doi:10.1016/j.neuroimage.2003.08.023
- Awasthi, B., Friedman, J., & Williams M. A. (2011a). Processing of low spatial frequency faces at periphery in choice reaching tasks. *Neuropsychologia, 49*, 2136-2141. doi:10.1016/j.neuropsychologia.2011.03.003
- Awasthi, B., Friedman, J., & Williams, M. A. (2011b). Faster, stronger, lateralized: low spatial frequency information supports face processing. *Neuropsychologia, 49*, 3583-3590. doi:10.1016/j.neuropsychologia.2011.08.027
- Awasthi, B., Sowman, P. F., Friedman, J., & Williams, M. A. 2013. Distinct spatial scale sensitivities for early categorization of faces and places; neuromagnetic and behavioural findings. *Frontiers in Human Neuroscience, 7*, 1-11. doi:10.3389/fnhum.2013.00091
- Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology, 3*, 87-103. doi:10.1080/09541449108406221
- Bayliss, A. P., di Pellegrino, G., & Tipper, S. P. (2004). Orienting of attention via observed eye gaze is head-centred. *Cognition, 94*, B1-B10. doi:10.1016/j.cognition.2004.05.002.

- Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, *50*, 2577-2587. doi:10.1016/j.visres.2010.08.016
- Bindemann, M., Attard, J., Leach, A., & Johnston, R. A. (2013). The effect of image pixilation on unfamiliar-face matching. *Applied Cognitive Psychology*, *27*, 707-717. doi:10.1002/acp.2970
- Bindemann, M., & Burton, A. M. (2008). Attention to upside-down faces: An exception to the inversion effect. *Vision Research*, *48*, 2555-2561. doi:10.1016/j.visres.2008.09.001
- Bindemann, M., & Burton, A. M. (2009). The role of colour in human face detection. *Cognitive Science*, *33*, 1144-1156. doi:10.1111/j.1551-6709.2009.01035.x
- Bindemann, M., Burton, A. M., Leuthold, H., & Schweinberger, S. R. (2008). Brain potential correlates of face recognition: Geometric distortions and the N250r brain response to stimulus repetition. *Psychophysiology*, *45*, 535-544. doi:10.1111/j.1469-8986.2008.00663.x
- Bindemann, M., & Lewis, M. B. (2013). Face detection differs from categorization: Evidence from visual search in natural scenes. *Psychonomic Bulletin & Review*, *20*, 1140-1145. doi:10.3758/s13423-013-0445-9
- Bindemann, M., Scheepers, C., & Burton, A. M. (2009). Viewpoint and centre of gravity affect eye movements to human faces. *Journal of Vision*, *9*, 1-16. doi:10.1167/9.2.7
- Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: Do faces pop out? *Perception*, *26*, 1555-1570. doi:10.1068/p261555

- Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identity of faces. *Perception, 23*, 803-822.
doi:10.1068/p230803
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327. doi:10.1111/j.2044-8295.1986.tb02199.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews, 36*, 96-107. Retrieved from <http://cns-alumni.bu.edu/~yazdan/pdf/Bullier01.pdf>
- Burton, M. (2015). Averages: Recognising faces across realistic variation. Retrieve from <http://homepages.abdn.ac.uk/m.burton/pages/averages.html>
- Burton, A. M., & Bindemann, M. (2009). The role of view in human face detection. *Vision Research, 49*, 2026-2036. doi:10.1016/j.visres.2009.05.012.
- Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81*, 361-380. doi:10.1111/j.2044-8295.1990.tb02367.x
- Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology, 51*, 256-284. doi:10.1016/j.cogpsych.2005.06.003
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods, 42*, 286-291. doi:10.3758/BRM.42.1.286
- Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research, 41*, 1179-1208. doi:10.1016/S0042-6989(01)00002-5

- Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*, 641-651.
doi.org/10.1038/nrn1724
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of grating. *The Journal of Physiology*, *197*, 551-566.
doi:10.1113/jphysiol.1968.sp00857
- Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*, 985-994.
doi:10.1068/p3335
- Collin, C. A., Liu, C. H., Troje, N. F., McMullen, P. A., & Chaudhuri, A. (2004). Face recognition is affected by similarity in spatial frequency range to a greater degree than within-category object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 975-987.
doi:10.1037/0096-1523.30.5.975
- Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception*, *23*, 129-146.
doi:10.1068/p230129
- Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics*, *5*, 602-612.
doi:10.3758/BF03213093
- Crouzet, S. M., Kirchner, H. K., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*, 1-17. doi:10.1167/10.4.16
- Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology*, *342*, 1-9. doi:10.3389/fpsyg.2011.00342

- Cutting, J. E. (1987). Rigidity in cinema scenes from the front row, side aisle. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 323-334.
doi:10.1037/0096-1523.13.3.323
- Davies, G., Ellis, H., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, *63*, 180-187.
doi:10.1037/0021-9010.63.2.180
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Shared attention and the social brain: Gaze perception triggers automatic visuospatial orienting in adults. *Visual Cognition*, *6*, 509-540. Retrieved from <http://wexler.free.fr/library/files/driver>
- Duchowski, A. T., Cournia, N., & Murphy, H. (2004). Gaze-contingent displays: A review. *CyberPsychology & Behavior*, *7*, 621-634. Retrieved from <http://andrewd.ces.clemson.edu/gcd/adc04.pdf>
- Farkas, L. G., Katic, M. J., Forrest, C. R., Alt, K. W., Bagic, I., Baltadjiev, G., ...Yahia, E. (2005). International anthropometric study of facial morphology in various ethnic groups/races. *Journal of Craniofacial Surgery*, *16*, 615-646.
doi:10.1097/01.scs.0000171847.58031.9e.
- Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, *12*, 195-201. doi:10.1068/p120195n
- Fischer, T., Graupner, S. T., Velichkovsky, B. M., & Pannasch, S. (2013). Attentional dynamics during free picture viewing: Evidence from oculomotor behaviour and electrocortical activity. *Frontiers in Systems Neuroscience*, *7*, 1-8.
doi:10.3389/fnsys.2013.00017

- Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37*, 571-583. doi:10.1068/p5705
- Frey, A., Ionescu, G., Lemaire, B., Lopez-Orozco, F., Baccino, T., & Guerin-Dugue, A. (2013). Decision-making in information seeking on texts: An eye-fixation related potentials in investigation. *Frontiers in Systems Neuroscience, 7*, 1-22. doi:10.3389/fnsys.2013.00039
- Garrido, C., Duchaine, B., & Nakayama, K. (2008). Face detection in normal and prosopagnosic individuals. *Journal of Neuropsychology, 2*, 119-140. doi:10.1348/174866407X246843
- George, N., Jemel, B., Fiori, N., Chaby, L., & Renault, B. (2005). Electrophysiological correlates of facial decision: Insights from upright and upside-down Mooney-face perception. *Cognitive Brain Research, 24*, 663-673. doi:10.1016/j.cogbrainres.2005.03.017
- Gilad, S., Meng, M., & Sinha, P. (2009). Role of ordinal contrast relationships in face encoding. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 5353-5358. doi:10.1073/pnas.0812396106.
- Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research, 16*, 416-424. doi:10.1016/S0926-6410(03)00056-9
- Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception, 34*, 77-86. doi:10.1068/p5370

- 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, 313-332. doi:10.1016/S0364-0213(03)00002-8
- 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, 1023-1039. doi:10.1037/0096-1523.32.4.1023
- Halit, H., de Haan, M., Schyns, P. G., & Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, 1117, 154-161. doi:10.1016/j.brainres.2006.07.059
- Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: Effects of two dimensional filtered noise. *Science*, 180, 1194-1197. doi:10.1126/science.180.4091.1194
- Henderson, J. M., Luke, S. G., Schmidt, J., & Richards, J. E. (2013). Co-registration of eye movements and event-related potentials in connected-text paragraph reading. *Frontiers in Systems Neuroscience*, 7, 1-13. doi:10.3389/fnsys.2013.00028
- Hershler, O., Golan, T., Bentin, S., & Hochstein, S. (2010). The wide window of face detection. *Journal of Vision*, 21, 1-14. doi:10.1167/10.10.21.
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, 45, 1707-1724. doi:10.1016/j.visres.2004.12.021

- Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception, 31*, 1221-1240.
doi:10.1068/p3252
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cats striate cortex. *Journal of Physiology, 148*, 574-591.
doi:10.1113/jphysiol.1959.sp006308
- Jenkins, R. (2007). The lighter side of gaze perception. *Perception, 36*, 1266-1268.
doi.10.1068/p5745
- Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science, 319*, 435. doi:10.1126/science.1149656
- Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society of London. Series B: Biological sciences, 366*, 1671-1683. doi:10.1098/rstb.2010.0379
- Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience, 6*, 766-774. doi:10.1038/nrn1766
- Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*, 1-19.
doi:10.1016/0010-0277(91)90045-6
- Johnston, R. A., & Bindemann, M. (2013). Introduction to forensic face matching. *Applied Cognitive Psychology, 27*, 697-699. doi:10.1002/acp.2963
- Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces - the role of shape from shading and pigmentation cues. *Perception, 25*, 37-52. doi:10.1068/p250037

- Kingdom, F. A., Field, D. J., & Olmos, A. (2007). Does spatial invariance result from insensitivity to changes? *Journal of Vision*, 7, 1-13. doi:10.1167/7.14.11.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46, 1762-1776. doi:10.1016/j.visres.2005.10.002
- Kuehn, S. M., & Jolicoeur, P. (1994). Impact of quality of the image, orientation and similarity of the stimuli on visual search for faces. *Perception*, 23, 95-122. doi:10.1068/p230095
- Lewis, M. B., & Edmonds, A. J. (2003). Face detection: Mapping human performance. *Perception*, 32, 903-920. doi:10.1068/p5007
- Lewis, M. B., & Edmonds, A. J. (2005). Searching for faces in scrambled scenes. *Visual Cognition*, 12, 1309-1336. doi:10.1080/13506280444000535
- Lewis, M. B., & Ellis, H. D. (2003). How we detect a face: A survey of psychological evidence. *International Journal of Imaging Systems and Technology*, 13, 3-7. doi:10.1002/ima.10040
- Livingstone, M. S., & Hubel, D. H. (1988). Segregation of form, color, movement and depth: Anatomy, physiology and perception. *Science*, 240, 740-749. Retrieved from <http://www.hms.harvard.edu/bss/neuro/bornlab/nb204/papers/livingstone-hubel-segregation-science1988.pdf>
- Lynch, J. J., Silveira, L. C., Perry, V. H., & Merigan, W. H. (1992). Visual effects of damage to P ganglion cells in macaques. *Visual Neuroscience*, 8, 575-583. doi:10.1017/S0952523800005678

- Macchi, C. V., Simion, F., & Umiltà, C. (2001). Face preference at birth: the role of an orienting mechanism. *Developmental Science*, *4*, 101-108 doi:10.1111/1467-7687.00154
- Marr, D., & Hildreth, E. C. (1980). Theory of edge detection. *Proceeding of the royal society of London B*, *207*, 187-217. doi:10.1098/rspb.1980.0020
- Marshall, J. A., Burbeck, C. A., Ariely, J. P., Rolland, J. P., & Martin, K. E. (1996). Occlusion edge blur: A cue to relative visual depth. *Journal of the Optical Society of America A*, *13*, 681-688. doi:10.1364/JOSAA.13.000681
- Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, *355*, 344-346. doi:10.1038/355344a0
- Morris, J. S., de Gelder, B., Weiskrantz, L., & Dolan, R. J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain*, *124*, 1241-1252. doi:10.1093/brain/124.6.1241
- Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, *8*, 454-469. doi:10.3758/BF03196180
- Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, *39*, 3824-3833. doi:10.1016/S0042-6989(99)00096-6
- Nestor, A., Vettel, J. M., & Tarr, M. J. (2013). An application of noise-based image classification to BOLD Responses. *Human Brain Mapping*, *34*, 3101-3115. doi:10.1002/hbm.22128

- Nguyen, M. N., Hori, E., Matsumoto, J., Tran, A. H., Ono, T., & Nishijo, H. (2013). Neuronal responses to face-like stimuli in the monkey pulvinar. *European Journal of Neuroscience*, *37*, 35-51. doi:10.1111/ejn.12020
- Nguyen, M. N., Matsumoto, J., Hori, E., Maior, R. S., Tomaz, C., Tran, A. H., ... Nishijo, H. (2014). Neuronal responses to face-like and facial stimuli in the monkey superior colliculus. *Frontier in Behavioural Neuroscience*, *8*, 8-85. doi: 10.3389/fnbeh.2014.00085
- Norman, J., & Ehrlich, S. (1987). Spatial frequency filtering and target identification. *Vision Research*, *27*, 87-96. doi:10.1016/0042-6989(87)90145-3
- Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception*, *22*, 1287-1298. doi:10.1068/p221287
- Ojanpää, H., & Näsänen, R. (2003). Utilisation of spatial frequency information in face search. *Vision Research*, *43*, 2505-2515. doi:10.1016/S0042-6989(03)00459-0
- Olk, B., & Garay-Vado, A. M. (2011). Attention to faces: Effects of face inversion. *Vision Research*, *51*, 1659-1666. doi:10.1016/j.visres.2011.05.007
- Parker, D. M., & Costen, N. P. (1999). One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face processing. *Current Psychology: Development, Learning, Personality, Social*, *18*, 118-127. doi:10.1007/s12144-999-1021-3
- Purcell, D. G., & Stewart, A. L. (1986). The face-detection effect. *Bulletin of the Psychonomic Society*, *24*, 118-120. Retrieved from

http://download.springer.com/static/pdf/640/art%253A10.3758%252FBF03330521.pdf?auth66=1423133284_3d4ae272c8ba33035498c648adfcf5a6&ext=.pdf

Purcell, D. G., & Stewart, A. L. (1988). The face-detection effect: Configuration enhances perception. *Perception & Psychophysics*, *43*, 355-366. Retrieved from http://download.springer.com/static/pdf/80/art%253A10.3758%252FBF03208806.pdf?auth66=1423068003_bfe061b7c01adf7c6e81e902a52c9b6a&ext=.pdf

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science*, *12*, 94-99. doi:10.1111/1467-9280.00317

Rolls, E. T., Baylis, G. C., & Leonard, C. M. (1985). Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. *Vision Research*, *25*, 1021-1035. doi:10.1016/0042-6989(85)90091-4

Sagi, D., & Julesz, B. (1986). Short range limitation on detection of feature differences. *Spatial Vision*, *12*, 1-10. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.367.654&rep=rep1&type=pdf>

Sandford, A., & Burton, A. M. (2014). Tolerance for distorted faces: Challenges to a configural processing account of familiar face recognition. *Cognition*, *132*, 262-268. doi:10.1016/j.cognition.2014.04.005.

- Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features!
Understanding recognition from the use of visual information. *Psychological Science, 13*, 402-409. doi:10.1111/1467-9280.0047
- Schyns, P. G., & Oliva, A. (1997). Flexible, diagnostically-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception, 26*, 1027-1038.
- Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition, 69*, 243-265. doi:10.1016/s0010-0277(98)00069-9
- Shapley, R., & Hawken, M. (2002). Neural mechanisms for color perception in the primary visual cortex. *Current Opinion in Neurobiology, 12*, 426-432.
doi:10.1016/S0959-4388(02)00349-5
- Simion, F., Farroni, T., Cassia, V., Turati, C., & Barba, B. (2002). Newborns' local processing in schematic facelike configurations. *British Journal of Developmental Psychology, 14*, 257-273. doi:10.1348/026151002760390800
- Simion, F., Macchi, V., Turati, C., & Valenza, E. (2003). Non-specific perceptual biases at the origin of face processing. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood* (pp.13-26). New York: Nova Science Publishers, Inc.
- Sinha, P. (2002). Qualitative representations for recognition [PDF document].
Lectures notes in computer science. Retrieve from
http://link.springer.com/chapter/10.1007%2F3-540-36181-2_25#page-2

- Smith, W. S., & Tadmor, Y. (2013). Nonblurred regions show priority for gaze direction over spatial blur. *The Quarterly Journal of Experimental Psychology*, *66*, 927-945. doi:10.1080/17470218.2012.722659.
- Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869-873. doi:10.1038/23703
- Tieger, T., & Ganz, L. (1979). Recognition of faces in the presence of two-dimensional sinusoidal masks. *Perception & Psychophysics*, *26*, 163-167. doi:10.3758/BF03208310
- Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97-136. doi:10.1016/0010-0285(80)90005-5
- Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, *114*, 285-310. doi:10.1037/0096-3445.114.3.285
- Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, *31*, 411-437. doi:10.1146/annurev.neuro.30.051606.094238
- Turati, C., Simion, F., Milani, I., & Umiltà, C. (2002). Newborns' preference for faces: What is crucial? *Developmental Psychology*, *38*, 875-888. doi:10.1037/0012-1649.38.6.875
- Ullman S, Vidal-Naquet M, & Sali E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*, 682-687. doi:10.1038/nn870

- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*, *43A*, 161-204. doi:10.1080/14640749108400966
- Valentine, T., & Bruce, V. (1986). The effect of distinctiveness in recognizing and classifying faces. *Perception*, *15*, 525-533. doi:10.1068/p150525
- Viola, P., & Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*, 137-154. doi:10.1023/B:VISI.0000013087.49260.fb
- Vuilleumier, P. (2000). Faces call for attention: Evidence from patients with visual extinction. *Neuropsychologia*, *38*, 693-700. doi:10.1016/S0028-3932(99)00107-4
- Westheimer, G. (2001). The Fourier theory of vision. *Perception*, *30*, 531-541. doi:10.1068/p3193
- Wolfe, J. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202-238. doi:10.3758/BF03200774
- Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception*, *31*, 995-1003. doi:10.1068/p33

Appendix: FACE-ABSENT SCENES USED IN EXPERIMENT 1



An example of face-absent scene for original condition.



An example of face-absent scene for LSF condition.



An example of face-absent scene for MSF condition.



An example of face-absent scene for HSF condition.