THE EFFECTS OF STIMULUS INFORMATION AND ORIENTATION ON FACE PROCESSING

by

Carl M. Gaspar, B.A., M.A.

A Thesis Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Department of Psychology, Neuroscience and Behaviour
McMaster University

Your file Votre référence
ISBN: 978-0-494-36084-2
Our file Notre référence
ISBN: 978-0-494-36084-2

Effects of Information and Orientation on Face Processing

DOCTOR OF PHILOSOPHY (2007) McMaster University
(Psychology, Neuroscience & Behaviour) Hamilton, Ontario, Canada

TITLE: The effects of stimulus information and orientation on face processing
AUTHOR: Carl M. Gaspar
SUPERVISORS: Patrick J. Bennett & Allison B. Sekuler
NUMBER OF PAGES: ix, 108

Abstract

The effects of stimulus information and orientation on face processing

Carl M. Gaspar, B.A., M.A.
Doctor of Philosophy
Department of Psychology, Neuroscience and Behaviour
McMaster University
2007

The face inversion effect refers to the reduction in the accuracy and speed of face identification when faces are shown upside-down, or "inverted". The face inversion effect is of great importance to psychologists studying face identification because it is thought to provide important clues about the information that humans use to identify faces. The common approach to studying the relationship between the face inversion effect and facial information assumes that upright and inverted face identification rely on qualitatively different processes. Contrary to this assumption, Sekuler, Gaspar, Gold and Bennett (2004) speculated that inverted faces are identified less accurately because of a subtle change in how information is used in the same, local regions of the eyes and eyebrows. Chapters 2 and 3 examined this hypothesis. Chapter 2 provides evidence in support of the speculation of Sekuler et al. and against the notion that upright and inverted face identification rely on qualitatively different processing mechanisms. Chapter 3 compared the spatial frequency tuning of upright and inverted face identification and found that the same narrow band of spatial frequencies were used for both types of stimuli. Therefore, the information processing strategy used for both upright and inverted face identification is similar in both the regions of the face used and the spatial frequency components that are used. If the upright and inverted face identification differ quantitatively, rather than qualitatively, then the common approach to uncovering critical facial information by studying the face inversion effect is not valid. Based on studies that have applied this approach, many researchers have concluded that humans rely critically on differences in the spacing among facial features, or relational cues, in order to identify faces. Experiments in Chapter 4 examine relational cues that could be useful for face identification in natural contexts by measuring thresholds for the discrimination of both interpupillary distance (IPD) and nose-to-mouth distance (NMD). Our findings show that people's spatial resolution for these cues is not fine enough to resolve the variation in IPD and NMD that exists in a large population of faces.

Acknowledgements

I would like to thank Patrick Bennett and Allison Sekuler for being fantastic mentors and great people; Donna Waxman for organizing my lab life and creating a jovial atmosphere; my parents and sisters for being patient and very supportive; and the entire vision lab, both past and present members, for not hitting me over the head with my tupperware.

Chapters 2 and 3 are based on papers submitted to the journal Vision Research (Gaspar, Bennett, & Sekuler, 2007 a&b). In both cases, I was involved in all aspects of the research: generating the experimental hypotheses, collecting and analyzing the data, and preparing the manuscript.

References

Gaspar, C.M., Bennett, P.J., & Sekuler, A.B. (2007a). The effects of face inversion and contrast reversal on efficiency and internal noise. Submitted to *Vision Research.*

Gaspar, C.M., Bennett, P.J., & Sekuler, A.B. (2007b). Spatial frequency tuning of upright and inverted face identification. Submitted to *Vision Research.*

Contents

Chapter 1

General Introduction

A fundamental issue in object recognition is how human observers recognize faces. Faces are a highly important visual signal for our social and interpersonal lives (Baron-Cohen & Cross, 1992; Baron-Cohen, Riviere, Cross, Fukushima, Bryant, Sotillo, Hadwin & French, 1996; Bruce, Burton, Hanna, Healey, Mason, Coombes, Fright & Linney, 1993; Ekman, 1984; George & Hole, 1995; Maloney & Dal Martello, 2006; Massaro & Stork, 1998; Perrett, Lee, Penton-Voak, Rowland, Yoshikawa, Burt, Henzi, Castles & Akamatsu, 1998). Among the various classes of objects we might encounter in our natural visual environment, the task of recognizing individual faces may be the most frequently performed recognition task. Therefore, face recognition may be one of the most meaningful and well-practiced of tasks in object recognition. However, the computational mechanism that underlies human face recognition is still poorly understood. One important challenge for researchers in this field is to characterize the most critical elements of information in a face that human observers depend on for person identification.

In psychology, a common theoretical distinction is made between the potential difference in how facial features are encoded compared to the spatial arrangement of those features (Collishaw & Hole, 2000; Diamond & Carey, 1986; Leder & Bruce, 2000; Leder, Candrian, Huber & Bruce, 2001; Maurer, Le Grand & Mondloch, 2002; Rhodes, 1988; Searcy & Bartlett, 1996). Facial features commonly refer to the eyes, the nose, the mouth, the ears, and any other local region of the face with a characteristic shape. The spatial arrangement of facial features refers specifically to the exact distances among features, and are sometimes referred to as configural or relational cues (Leder & Bruce, 2000). In other words, psychologists have, for the most part, concerned themselves with how human observers use those aspects of a face that are well described by everyday language. However, this description of facial features and configural cues may not necessarily map onto how faces are coded in the human brain. There are numerous alternative approaches to the features-versus-relations distinction that can be found in the field of automated face recognition (Zhao, Chellappa, Phillips & Rosenfeld, 2003). Some of these automated systems (Lades et al., 1993; Sirovich & Kirby, 1987) have been used to successfully model how human observers recognize faces (Burton et al., 2005; Burton et al., 2001; Yue, Tjan & Biederman, 2006). However, many of those physical aspects do not not correspond to anything we can easily label with common terms; for example, eigenfaces (Sirovich & Kirby, 1987) and Gabor-jets (Lades et al., 1993). None of this should be surprising to computer scientists or psychologists. The human face is a complex spatial pattern and naturally lends itself to numerous forms of representation, any of which might be a candidate for modelling human face recognition. Those representations that appeal more to intuition, like the features and relational cues, are simply one of the many representations that a theorist may evaluate.

However, even if one were to decide upon a particular kind of representation, the complexity of the human face poses additional problems. For example, suppose that relational cues are in fact critical for face recognition. How then do we decide on which specific feature relations are most important for face recognition? On the frontal-parallel view of the human face alone, the facial anthropometrist Leslie G. Farkas has listed 38 unique landmarks that anthropometrists and surgeons can use to reliably map out the normal geometry of distances in a population of faces (Farkas, 1981). Considering only the distances between pairs of these facial landmarks, there are already 703 distances to consider. If some psychologists are correct in supposing that configural cues are important for face recognition, which among these distances are most important for human observers? This tedious task has only been partly undertaken (Haig, 1984; Sergent, 1984), and one can easily see why - the possibilities are too many. In the spatial frequency domain, faces are no less complicated; facial images have broad amplitude spectra and, depending on the conditions of a face identification task, information that is useful for identification may be distributed across the entire spectra (Gold, Bennett & Sekuler, 1999; Näsänen, 1999). In conclusion, the complexity of the human face as a spatial pattern poses a problem for researchers who wish to uncover those characteristics of the face that are most important for recognition.

Despite the challenges involved in specifying the critical aspects of facial images, there is an optimism among many psychologists and cognitive scientists that we have already learned a great deal about facial information from the results of studies on the face inversion effect. The face inversion effect refers to the reduction in identification accuracy and speed that occurs when faces are viewed upside-down (inverted in orientation) compared to when they are viewed upright. The face inversion effect is a highly robust effect that can be observed in a diverse number of face identification paradigms, and across a wide range of face images (Martelli, Majaj & Pelli, 2005; Valentine, 1988). The face inversion effect has been, and still is, of great interest to researchers in the field of object recognition. One reason why the face inversion effect has received so much attention is because many researchers believe that this effect can be exploited to provide us with clues about the kinds of facial characteristics we commonly use to identify faces (Barton, Keenan & Bass, 2001; Collishaw & Hole, 2000; Diamond & Carey, 1986; Farah, Tanaka & Drain, 1995; Farah, Wilson, Drain & Tanaka, 1995; Freire, Lee & Symons, 2000; Gauthier, Williams, Tarr & Tanaka, 1998; Kemp, McManus & Pigott, 1990; Leder & Bruce, 2000; Leder et al., 2001; Maurer et al., 2002; Rhodes, 1988; Robbins & McKone, 2006; Searcy & Bartlett, 1996; Yin, 1969; Young, Hellawell & Hay, 1987). Very briefly, this approach assumes that qualitatively different perceptual mechanisms support the recognition of upright and inverted faces. The mechanism that supports upright face recognition exclusively, while disrupted by stimulus inversion, is thought use information that is critical for normal face recognition. Based on these assumptions, therefore, some researchers have concluded that we can uncover the elements of information that are most critical for normal face recognition by measuring the magnitude of the inversion effect in face identification tasks designed so that the observer is forced to use specific kinds of information. Facial information associated with a strong inversion effect is then taken to be under the exclusive domain of the mechanism underlying upright face processing and, therefore, believed to be critical for normal face recognition.

To illustrate the common approach to understanding the inversion effect, consider the study by Leder and Bruce (2000). Leder and Bruce designed a relational set of faces that shared the same facial features but differed in how those features were spaced. They also designed a set of parts-only faces that shared the same spacing among features, but differed by the appearance of individual facial features. Identification accuracy was measured for both sets of faces, shown both in their normal upright orientation and also inverted. The inversion effect was measured as the decrement in identification accuracy in inverted conditions compared to upright conditions. Leder and Bruce obtained larger inversion effects for the relational set of faces compared to the parts-only set of faces, and they concluded that "...the inversion effect in faces arises from the processing of critical relational cues from upright faces" (page 525). Within the same study, Leder and Bruce conduct similar experiments and, from the results of those experiments, they state that: "The critical information that is used in face recognition and that is disrupted by inversion consists of relations between single features" (page 534). Based in part on studies that have applied this approach to understanding face recognition, many researchers have concluded that so-called configural or relational cues, the variations in the exact spacing among facial features (like the eyes, and mouth), are critical for normal face recognition (Diamond & Carey, 1986; Gauthier et al., 1998; Kemp et al., 1990; Le Grand, Mondloch, Maurer & Brent, 2001; Le Grand, Mondloch, Maurer & Brent, 2003; Leder & Bruce, 2000; Leder et al., 2001; Maurer et al., 2002; Rhodes, 1988; Searcy & Bartlett, 1996).

There are two major problems with research that has taken the approach just described. First, there is no clear evidence that upright and inverted face recognition rely on qualitatively different processes. Almost two decades ago, Valentine (1988) published a review of the face inversion effect that pointed out this very problem. However, it was not until recently that experimental research directly examined the idea that upright and inverted face identification differ in a qualitative manner (Riesenhuber, Jarudi, Gilad & Sinha, 2004; Sekuler, Gaspar, Gold & Bennett, 2004). Sekuler et al. (2004) used a psychophysical technique called reverse correlation to directly visualize regions of the face that observers use to identify both upright and inverted face recognition. Their results were surprising because they showed that very similar, local regions of the face — around the eyes and eyebrows — were used for both upright and inverted face recognition. Despite the close similarity in the information used to identify upright and inverted faces, there also appeared to be very subtle, quantitative differences in how the small regions around the eyes and eyebrows were sampled. The results of Sekuler et al. suggest that the difference between upright and inverted face processing is quantitative, rather than qualitative.

The common approach of using the face inversion effect to reveal the critical elements of facial identity suffers from a second problem. This approach fails to take into account one of the most fundamental constraints on any information-processing strategy that must succeed at face recognition: facial characteristics that can support accurate face identification must vary significantly across individuals in the population. There is no clear evidence that the relational cues available in real faces are variable enough to support face identification in a natural context.

Chapters 2 and 3 of this thesis describe studies that use different psychophysical techniques to characterize the information-processing strategy used during both upright and inverted face recognition. The results of these two chapters extend the results of Sekuler et al. (2004) by providing further evidence that is contrary to the notion that upright and inverted face recognition rely on qualitatively different processes. In fact, the results of these two chapters, together with those of Sekuler et al., suggest that the information-processing strategy used for face recognition is strikingly similar for upright and inverted faces. Therefore, the results of Chapters 2 and 3 call into question the common experimental approach that uses the inversion effect to uncover the characteristics of the face that are most critical for normal face recognition. However, this result does not necessarily mean that relational cues are unimportant for face recognition. Nonetheless, the importance of relational cues, or any facial cue for that matter, must be confirmed by determining if their variability in a population of faces is large enough to support face recognition. Chapter 4 therefore presents a novel approach that makes use of data from facial anthropometry to assess the ecological utility of configural cues. Our results demonstrate that two of the most commonly studied configural cues, interpupillary distance and nose-to-mouth distance, would not be particularly useful cues for face recognition in a natural context.

Part of this thesis questions how researchers have commonly used the inversion effect to further their understanding of face recognition. However, our results do not exclude the possibility that the inversion effect could be used to better understand normal face recognition in other ways. On the contrary, our studies demonstrate similarities between upright and inverted face processing that are striking in light of the fact that inverted faces are almost never encountered in one's natural environment. Chapter 5 explains that one way to resolve this issue is to suppose that the facial characteristics most critical for face recognition may, in fact, look similar when either upright or inverted. This speculation raises some interesting theoretical questions about the kind of knowledge that we gain during our extended experience performing pattern recognition tasks. Chapter 5 also discusses a very important phenomenon in the face recognition literature called the Thatcher Illusion (Thompson, 1980). This striking perceptual phenomenon is perhaps the experimental finding that is most often cited in support of a qualitative difference between upright and inverted face recognition. However, I suggest that the Thatcher Illusion is also consistent with the alternate framework suggested by this thesis, that information-processing for upright and inverted face recognition rely on similar information-processing strategies.

1.1 References

Baron-Cohen, S., & Cross, P. (1992). Reading the eyes: Evidence for the role of perception in the development of a theory of mind. Mind and Language, 6, 173-186.

Baron-Cohen, S., Riviere, A., Cross, P., Fukushima, M., Bryant, C., Sotillo, M., Hadwin, J., & French, D. (1996). Reading the mind in the face: A cross-cultural and developmental study. Visual Cognition, 3, 39-59.

Barton, J.J., Keenan, J.P., & Bass, T. (2001). Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. Br J Psychol, 92 Part 3, 527-549.

Bruce, V., Burton, A.M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces? Perception, 22, 131-152.

Burton, A.M., Jenkins, R., Hancock, P.J.B., & White, D. (2005). Robust representations for face recognition: The power of averages. Cognitive Psychology, 51 (3), 256-284.

Burton, A.M., Miller, P., Bruce, V., Hancock, P.J.B., & Henderson, Z. (2001). Human and automatic face recognition: a comparison across image formats. Vision Research, 41 (24), 3185-3195.

Collishaw, S.M., & Hole, G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. Perception, 29 (8), 893-909.

Diamond, R., & Carey, S. (1986). Why Faces Are and Are Not Special - an Effect of Expertise. Journal of Experimental Psychology-General, 115 (2), 107-117.

Ekman, P. (1984). Expression and the nature of emotion. In: K.R. Scherer, & P. Ekman (Eds.), Approaches to Emotion (pp. 319-343). Hillsdale, N.J.: Lawrence Erlbaum.

Farah, M.J., Tanaka, J.W., & Drain, H.M. (1995). What Causes the Face Inversion Effect. Journal of Experimental Psychology-Human Perception and Performance, 21 (3), 628-634.

Farah, M.J., Wilson, K.D., Drain, H.M., & Tanaka, J.R. (1995). The Inverted Face Inversion Effect in Prosopagnosia - Evidence for Mandatory, Face-Specific Perceptual Mechanisms. Vision Research, 35 (14), 2089-2093.

Farkas, L.G. (1981). Anthropometry of the Head and Face in Medicine. (New York: Elsevier.

Freire, A., Lee, K., & Symons, L.A. (2000). The face-inversion effect as a deficit in the encoding of configural information: direct evidence. Perception, 29 (2), 159-170.

Gauthier, I., Williams, P., Tarr, M.J., & Tanaka, J. (1998). Training 'greeble' experts: a framework for studying expert object recognition processes. Vision Res, 38 (15-16), 2401-2428.

George, P.A., & Hole, G.J. (1995). Factors influencing the accuracy of age estimates of unfamiliar faces. Perception, 24, 1059-1073.

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

Haig, N.D. (1984). The effect of feature displacement on face recognition. Perception, 13, 505-512.

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. Perception, 19 (4), 531-543.

Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., Vandermalsburg, C., Wurtz, R.P., & Konen, W. (1993). Distortion Invariant Object Recognition in the Dynamic Link Architecture. Ieee Transactions on Computers, 42 (3), 300-311.

Le Grand, R., Mondloch, C.J., Maurer, D., & Brent, H.P. (2001). Neuroperception. Early visual experience and face processing. Nature, 410 (6831), 890.

Le Grand, R., Mondloch, C.J., Maurer, D., & Brent, H.P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. Nat Neurosci, 6 (10), 1108-1112.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 53 (2), 513-536.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. Perception, 30 (1), 73-83.

Maloney, L.T., & Dal Martello, M.F. (2006). Kin recognition and the perceived facial similarity of children. Journal of Vision, 10, 1047-1056.

Martelli, M., Majaj, N.J., & Pelli, D.G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. Journal of Vision, 5 (1), 58-70.

Massaro, D.W., & Stork, D.G. (1998). Speech recognition and sensory integration. American Scientist, 86, 236-244.

Maurer, D., Le Grand, R., & Mondloch, C.J. (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6 (6), 255-260.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. Vision Res, 39 (23), 3824-3833.

Perrett, D.I., Lee, K.J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D.M., Henzi, S.P., Castles, D.L., & Akamatsu, S. (1998). Effects of sexual dimporphism on facial attractiveness. Nature, 394, 884-887.

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. Perception, 17 (1), 43-63.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. Proceedings of the Royal Society of London Series B-Biological Sciences, 271, S448-S450.

Robbins, R., & McKone, E. (2006). No face-like processing for objects-of-expertise in three behavioural tasks. Cognition,

Searcy, J.H., & Bartlett, J.C. (1996). Inversion and processing of component and spatial-relational information in faces. J Exp Psychol Hum Percept Perform, 22 (4), 904-915.

Sekuler, A.B., Gaspar, C.M., Gold, J.M., & Bennett, P.J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. Current Biology, 14 (5), 391-396.

Sergent, J. (1984). Configural processing of faces in the left and the right cerebral hemispheres. J Exp Psychol Hum Percept Perform, 10 (4), 554-572.

Sirovich, L., & Kirby, M. (1987). Low-Dimensional Procedure for the Characterization of Human Faces. Journal of the Optical Society of America a-Optics Image Science and Vision, 4 (3), 519-524.

Thompson, P. (1980). Margaret Thatcher: a new illusion. Perception, 9 (4), 483-484.

Valentine, T. (1988). Upside-down Faces - a Review of the Effect of Inversion Upon Face Recognition. British Journal of Psychology, 79, 471-491.

Yin, R.K. (1969). Looking at upside-down faces. Journal of Experimental Psychology, 81, 141-145.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configurational information in face perception. Perception, 16 (6), 747-759.

Yue, X., Tjan, B.S., & Biederman, I. (2006). What makes faces special? Vision Res, 46 (22), 3802-3811.

Zhao, W., Chellappa, R., Phillips, P.J., & Rosenfeld, A. (2003). Face recognition: A literature survey. Acm Computing Surveys, 35 (4), 399-459.

Chapter 2

Efficiency and Internal Noise

What computations are performed to recognize a human face? To understand the neural system underlying face recognition, we need to describe both its capabilities and its limitations. The recognition of inverted (e.g., upside-down) and contrast-reversed faces is especially illuminating: Faces transformed in these ways contain all of the information that a normal face provides, and yet humans recognize the transformed faces less accurately and more slowly (Yin, 1969; Galper, 1970). The face inversion and contrast-reversal effects reflect constraints on human vision that are highly consistent across observers and paradigms (for reviews see Martelli, Majaj & Pelli, 2005 and Vuong, Peissig, Harrison, & Tarr, 2005).

The reliability of the face inversion effect has led some researchers to suggest that stimulus orientation affects the relative contributions of configural and feature-based processing to face perception (for a review see Maurer, Le Grand & Mondloch, 2002). However, Sekuler et al. (2004) took a different approach, and considered how upright and inverted face identification might differ if both are based on the application of local spatial filters (Ahumada, 2002; Murray, Bennett & Sekuler, 2002; Burgess, 1985). Sekuler et al. hypothesized that, even if inverted and upright face identification is based on the responses of local filters, there can be important differences in how regions across the face are sampled. Using the reverse correlation method, Sekuler et al. were able to map the influence of various pixels on responses in identification tasks that used upright and inverted faces. The resulting maps, called classification images, were surprisingly similar for upright and inverted faces, and showed that observers identified faces based on information conveyed by pixels in spatially limited regions around the eyes and eyebrows, regardless of face orientation. Nonetheless, there were quantifiable differences between the structure of classification images obtained with upright and inverted faces, and these differences were strongly correlated with the size of the face inversion effect found in different observers. Sekuler et al. concluded that although the spatial sampling strategies used for upright and inverted face identification differ only slightly, the differences are sufficient to account for the face inversion effect. Moreover, they found no evidence to support the idea that observers used fundamentally different processes to identify upright and inverted faces in their experimental conditions.

The conclusions drawn by Sekuler et al. (2004) assume that the levels of internal noise are equivalent when processing both upright and inverted faces, and that the effect of stimulus orientation on the structure within the classification images was due solely to differences in the efficiency with which information is extracted from stimuli (Murray, Bennett & Sekuler, 2005). However, it is plausible that internal noise levels are higher when processing inverted faces than when processing upright faces. For example, suppose that observers discriminate faces on the basis of responses of local spatial filters. Spatial jitter between the stimulus and a filter – which might arise as a result of eye movements – can introduce variation in the filter's response that is proportional to the stimulus contrast variance (Tyler & Chen, 2000). Therefore, if observers use similar sampling strategies to encode upright and inverted faces, but there is more spatial jitter in the inverted condition, then internal noise should be greater for inverted faces. Moreover, the idea that spatial jitter is greater for inverted faces is consistent with the report that patterns of fixa-

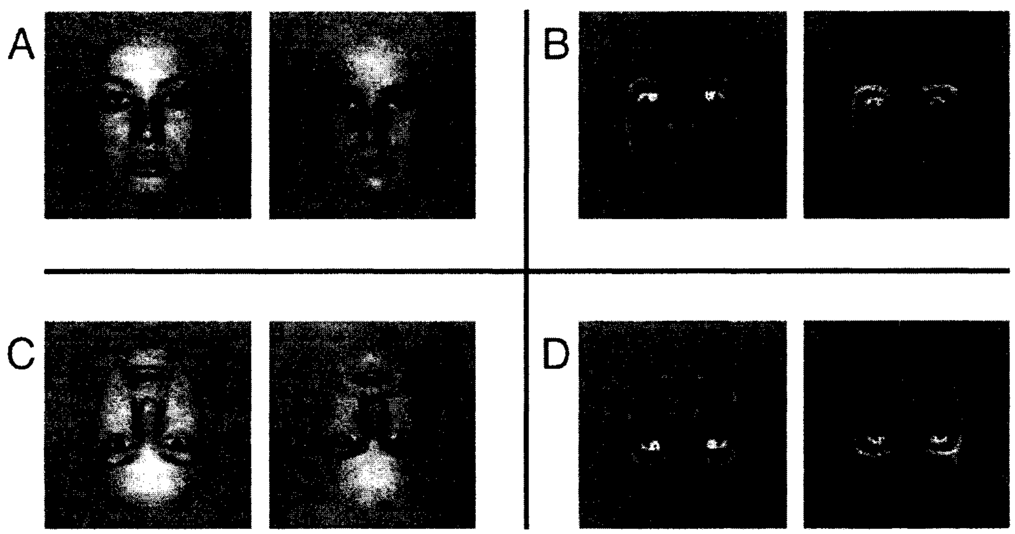


Figure 2.1: The pair of female faces used in the four main stimulus conditions: normal (A), reversed-contrast (B), inverted (C), and both reversed-contrast and inverted (D). An additional pair of male faces, not shown, was also used in each of the four conditions.

tions during face identification are more random for inverted faces (Barton, Radcliffe, Cherkasova, Edelman, & Intriligator, 2006). Fortunately, there is a psychophysical measure of an observer's response consistency that allows one to estimate the total quantity of internal noise associated with identification (Green, 1964; Burgess & Colborne, 1988; Gold, Bennett & Sekuler, 1999b; Gold, Sekuler & Bennett, 2004). When combined with the results of a noise-masking experiment, which provides an estimate of the relative efficiency of spatial sampling (Pelli & Farell, 1999), measures of response consistency can be used to estimate the separate constraints imposed by inefficient spatial sampling and elevated internal noise. In the current study, we combined noise-masking and response consistency methods to measure the internal noise and efficiency associated with identifying normal, inverted, and contrast-reversed faces. The inclusion of contrast-reversed faces allowed us to compare face inversion with another type of transformation on faces that makes them look less familiar while, at the same time, preserving the information that is available for identification.

2.0.1 The noise-masking technique

We measured contrast thresholds for two-alternative face identification both with and without the addition of display noise (see Figure 2.1). Because the face inversion and contrast-reversal effects are well-established results, we expected contrast thresholds in the absence of external noise to be higher for inverted and contrast-reversed faces (see Figure 2.2a). In other words, observers should need more contrast to identify these transformed faces accurately, compared to normal faces. As described by Pelli (1990), higher thresholds may reflect the contributions of two different types of constraints: a decrease in high-noise efficiency, or an increase in equivalent input noise. Pelli described a method to estimate high-noise efficiency and equivalent input noise by measuring contrast thresholds both with and without external noise added to the stimulus. In the current study, we applied this noise-masking method to the recognition of normal, inverted, and contrast-reversed faces, and also faces that were both inverted and contrast-reversed. By comparing the high-noise efficiency and equivalent input noise of normal face recognition to those that of our transformed faces, we characterized the constraints on performance associated with each stimulus transformation.

Across a wide range of psychophysical tasks, thresholds, expressed as squared rms contrast, c_{rms}^2, are a linear function of external noise variance, σ_n^2. (Pelli, 1998; Bennett, Sekuler & Ozin, 1999; Legge, Kersten & Burgess, 1987; Pelli & Farell, 1999; Raghavan, 1989; Tjan, Braje, Legge & Kersten, 1995; Dosher & Lu, 2000; Lu & Dosher, 1998; Lu, Liu & Dosher, 2000). Gold, Bennett and Sekuler (1999) and Gold, Sekuler and Bennett (2004) demonstrated that the function relating face discrimination thresholds to external noise, herein referred to as the noise-masking function, also is linear. Therefore, the noise-masking function for faces can be fully characterized by two independent parameters: σ_c^2 and k:

$$c_{rms}^2 = k(\sigma_n^2 + \sigma_c^2) \tag{2.1}$$

The parameter σ_c^2 , which is referred to as the equivalent input noise, is defined as the external noise variance that doubles threshold over a zero-noise baseline. The slope, k, is proportional to the effective signal-to-noise ratio at threshold, and is inversely proportional to an observer's high-noise efficiency (Pelli & Farell, 1999). A low value of k occurs when sampling efficiency is high and/or when contrast-dependent internal noise is low.

At one extreme, the face inversion or contrast-reversal effects could be fully characterized by an increase in σ_c^2. For example, inversion could increase contrast thresholds by the same amount across all levels of display noise. In this case, the noise-masking functions for inverted (or contrast-reversed) and normal faces would be the dotted and solid lines, respectively, in Figure 2.2b. Theoretically, this scenario is consistent with the hypothesis that the primary constraint on the perception of inverted faces is an increased amount of contrast-invariant internal noise. An increase in equivalent input noise is consistent with the idea that the representations of inverted (or contrast-reversed)

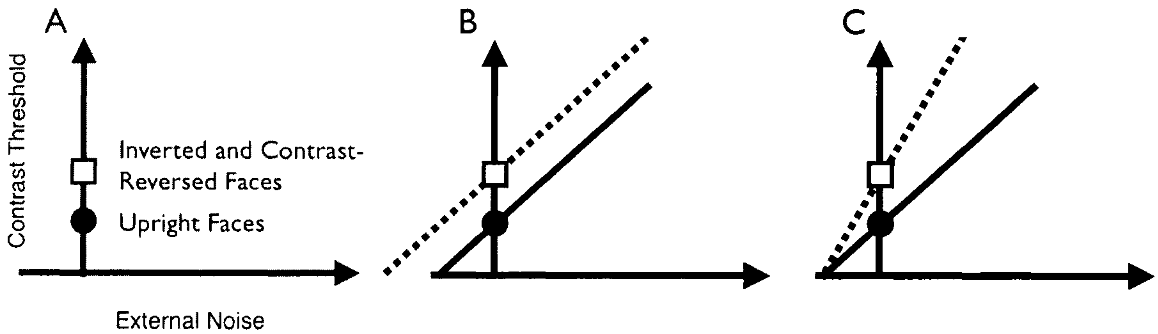


Figure 2.2: Contrast thresholds plotted as a function of the level of display noise on linear axes. Thresholds for identification without display noise (A) are expected to be higher for inverted and contrast-reversed faces (hollow square), relative to those for normal faces (solid circle). An increase in equivalent input noise results in thresholds rising by the same amount for all levels of external noise; e.g., the dotted line in B is a vertically shifted copy of the solid line. A decrease in calculation efficiency results in thresholds that rise faster with increasing amounts of external noise; e.g., the dotted line in C has a steeper slope than the solid line.

faces are corrupted by random variations in local contrast beyond those created by display noise. Many low-level factors may contribute to equivalent input noise (e.g., the quantum fluctuation of photons, variability in neuronal firing rate; for reviews see Pelli, 1990 and Raghavan, 1995). High-level factors may also contribute to increased equivalent input noise, even if variability remains constant at the level of individual neurons. For example, face inversion and contrast-reversal might result in a degradation of visual attention, which is known to modulate the contrast gain of individual neurons at relatively early stages of visual processing like V4 (for example, see Reynolds, Pasternak & Desimone, 2000). Because of the reduction in early signal gain, any neuronal noise introduced after the attention-dependent change in gain will have a greater influence on behavior, relative to stimulus-related activity. Psychophysically, this type of attention-modulated contrast-gain would manifest itself as an increase in contrast-invariant internal noise. Any other type of early signal attenuation that occurs before a late noise source can also result in increased contrast-invariant internal noise.

At the other extreme, either the face inversion or contrast-reversal effect could be fully characterized by an increase in the slope of the noise-masking function, k (Figure 2.2c). Decreased high-noise efficiency may be caused by an inefficient information-sampling strategy. In our task, the ideal strategy is to cross-correlate the stimulus with templates that match each of the two possible targets, and select the item that yields the highest cross-correlation. An equivalent strategy is to compute a linear combination of contrast values in the stimulus and then select a response based on whether the sum is greater or less than some criterion. In our task, the optimal set of contrast weights corresponds to the difference between the two images, which can be thought of as a map of the information available for discrimination (Figure 2.3a). An observer who completely ignores any region of the face containing information is using an inefficient sampling strategy. For example, the weighting schemes depicted in Figures 2.3b and 2.3c are suboptimal, and therefore would lead to high values of k. However, the strategy depicted in Figure 2.3c considers pixels that denote larger differences between the targets and, therefore, is more efficient than the strategy in Figure 2.3b, leading to a lower value of k. The classification image results of Sekuler et al. (2004) suggest that observers performing both upright and inverted face identification rely on local spatial filters to perform the task, and that differences in the quality of these filters are responsible for the face inversion effect. However, their results also demonstrate that the differences between the spatial filters used for upright and inverted faces are subtle, and not like the gross change in the placement of local filters illustrated by the difference between Figures 2.3b and 2.3c.

2.0.2 Response consistency

The interpretation of the noise-masking function slope as the inverse of sampling efficiency is complicated by the fact that the visual system contains contrast-dependent (multiplicative) noise (Burgess & Colborne, 1988). Indeed, it can be shown that, in the context of our framework, a higher value of k

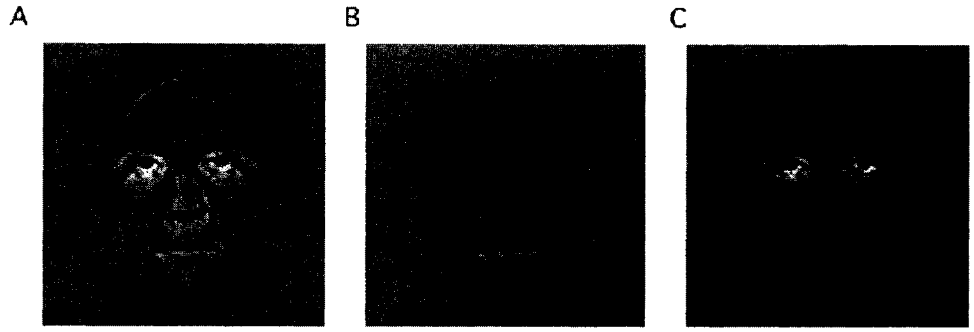


Figure 2.3: The ideal template in A is the difference between the two target images in one version of our recognition task (upright, male faces with normal contrast polarity). Pixels that are closer to black or to white, rather than grey, are highly informative. The linear templates shown in B and C are suboptimal because they fail to consider all pixels. The linear template in C is more efficient than the one in B because it considers more informative regions. Face orientation and contrast-polarity have no effect on how information is spread across the face. Face gender does have an effect; separate ideal templates exist for the pair of male and female faces we used.

could be due either to decreased sampling efficiency or to increased contrast-dependent noise (Gold et al., 1999b). To disambiguate the results, we use the so-called "double-pass technique," which measures the percent agreement between the responses to two identical sequences of stimuli-plus-noise combinations (Green, 1964; Burgess & Colborne, 1988; Gold et al., 1999b; Gold et al., 2004). The way in which percent correct and percent agreement co-vary is related to the internal:external noise ratio at the level of the decision variable (Green, 1964; Burgess & Colborne, 1988), as illustrated in Figure 2.4. The hollow squares depict hypothetical data from a task in which the internal:external noise ratio is relatively high: the hollow squares fall along a line that has a shallow slope. In contrast, the filled circles depict hypothetical data from a task in which the internal:external noise ratio is relatively low: in this case, the filled circles fall along a line with a steep slope. If inversion or contrast-reversal impair face recognition by increasing contrast-dependent internal noise, then the internal:external noise ratio should be higher in those conditions, and the line relating response accuracy and consistency should be shallower in those conditions than in a condition used normal faces. If, on the other hand, inversion or contrast-reversal does not alter contrast-dependent noise, then the relation between response accuracy and consistency should be invariant across conditions. Such a result would mean that any observed differences in high-noise efficiency could not be attributed to changes in contrast-dependent noise.

2.1 Methods

2.1.1 Observers

Eight observers (6 female, 2 male; average age = 22 years) participated in the experiment. All participants, except CG, were naïve about the purpose of the experiment and unfamiliar with the faces that were used as stimuli. Three participants (CG, AC and WL) were experienced psychophysical observers. All participants had normal or corrected-to-normal Snellen acuity. One subject did not complete all eight conditions, and so was excluded from the statistical analyses.

2.1.2 Stimuli & Apparatus

Stimuli were generated by a Macintosh G3 computer, and displayed on an AppleColor High-Resolution RGB monitor (model M0401) using MATLAB 5.1 and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The monitor resolution was set to 640 x 480 pixels at a frame rate of 67 Hz (non-interlaced). The monitor calibration data were used to build a 1779-element look-up table (Tyler, Chan, Liu, McBride & Kontsevich, 1992). Customized computer software constructed the stimuli on each trial by selecting the appropriate luminance values from the calibrated lookup table and storing them in the display's 8-bit lookup table. This procedure enabled us to manipulate contrast with high resolution: for example, pixel contrast could be varied from -0.99 to +1.70 in steps of approximately 0.0015.

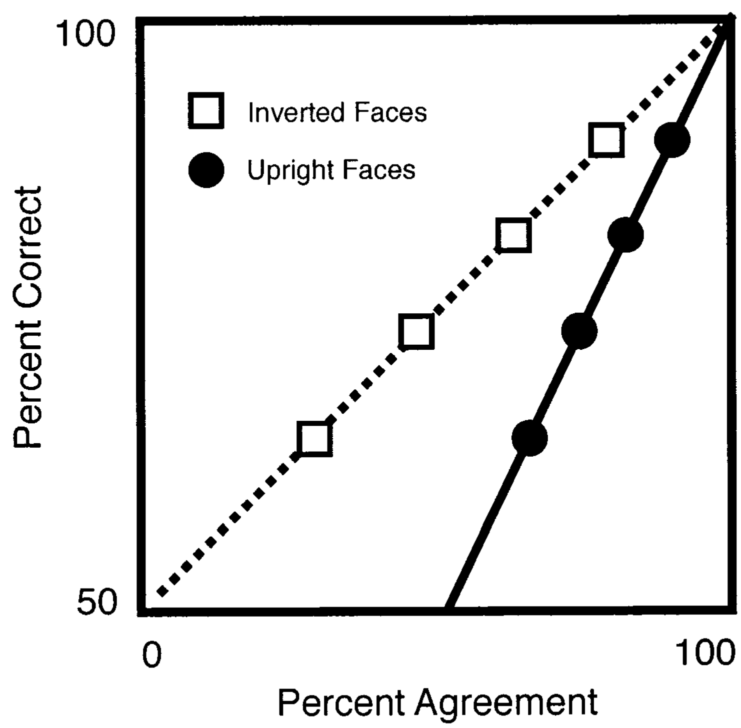


Figure 2.4: An example response-consistency plot based on hypothetical results. The Y-axis shows percentage of correct responses for a given level of image contrast. The X-axis shows the percentage of exactly repeated trial-pairs on which the observer gave the same response. Percent agreement is measured over all the trial-pairs for a given level of image contrast, giving us separate measures of agreement for each level of contrast that we test. The model data depicted with hollow squares is meant to reflect an observer with high internal noise, while the model data drawn in solid circles reflects low internal noise.

Face stimuli were based on digitized photographs of 4 faces (2 male and 2 female) cropped to an oval window, excluding areas showing the chin and hair, including the hairline (see Gold et al., 1999 for details). From the viewing distance of 1 meter, the height and width of each face subtended 2.0 and 1.4 deg respectively. Faces were centered within a 128 x 128 pixel square, and the amplitude spectrum of each image was set to the average spectrum across the original set of ten images. Faces were presented to observers in either a uniform field or embedded in white Gaussian noise (contrast variance σ_n^2= 0.0625) From the viewing distance of 1 meter, each face stimulus subtended 2.6 x 2.6 deg. Average luminance was 33 cd/m^2.

2.1.3 Procedure

Each participant performed two-alternative face identification in eight separate sessions, one for each possible combination of face gender, orientation, and contrast polarity. During each session, observers either discriminated the two male faces, or the two female faces; orientation and contrast polarity were held constant within each session as well. The order of conditions varied randomly among participants, except that sessions were grouped by face gender for two of the eight participants.

Each experimental session consisted of 2000 trials, lasting approximately one hour. Sessions began with a two-minute adaptation period, followed by a screen displaying the two face images to be discriminated during each test trial. Participants were instructed to carefully examine the faces during this time, and also during the selection period of each trial.

Trials consisted of the following sequence: First, a small fixation point (8 x 8 arc min) was displayed at the center of the screen for 1 s; across trials, the fixation point randomly changed from black to white. Participants were instructed to focus on this marker until stimulus presentation. The stimulus was then displayed in the center of the screen for 0.5 s. This test stimulus was one of the two possible faces (randomly selected on each trial), with external noise added where applicable. Finally, a selection screen appeared. This screen displayed both faces, without noise, at a contrast variance of 0.1. Participants were instructed to press a button corresponding to the face they believed was the test stimulus. Feedback was provided in the form of short audio beeps: a low tone for incorrect responses, and a high tone for correct responses. Participants were familiarized with this entire procedure in short demo sessions of approximately 20 trials before completing any experimental sessions.

Each session consisted of two types of randomly intermixed trials, in which the face stimuli were embedded in a high level of external noise (contrast variance = σ_n^2 = 0.0625) or presented on a uniform background (i.e., zero noise). For each trial type, stimulus contrast was manipulated by two interleaved staircases, one following the 2-down/1-up rule and the other following a 3-down/1-up rule. Each staircase continued for a total of 250 trials. The first 1000 trials, during which all four staircase runs were completed, were replicated exactly and in the same order for the second half of the session. All participants, except CG, were not aware that the second half of trials was a replicate of the first half.

2.1.4 Analyses

In each condition, the data from the two staircases were combined and a single psychometric function was estimated by computing the best-fitting Weibull function. Threshold was defined as the contrast variance needed to produce 71% correct responses. Equation 2.1 was fit to thresholds measured in zero-noise and high-noise conditions to derive k and σ_c^2 value for each participant in each condition. A bootstrap procedure of 1500 iterations was used to calculate the standard deviations of threshold, k, and σ_c^2. Monte Carlo simulations were used to estimate threshold of a simulated ideal observer at multiple non-zero noise-levels. Equation 2.1 was then fit to the resulting threshold-vs.-noise functions to estimate the values of the slope, k_{ideal}, for Male and Female stimuli. Values of k obtained from real observers with male and female faces were divided by the appropriate value of k_{ideal} and multiplied by 100 to obtain percent high-noise efficiency (Pelli & Farrell, 1999).

Our derivation of k and σ_c^2 assumes that thresholds are well fit by Equation 2.1. This assumption is reasonable for identification thresholds for normal faces, which have been shown to be consistent with Equation 2.1 (Gold et al., 1999b; 2004), however, we know of no previous work showing that discrimination thresholds for inverted or contrast-reversed faces conform to the predictions of the linear model. Therefore, we conducted a pilot study using the first author as the observer to determine if Equation 2.1 also provides a good fit to noise-masking functions obtained with inverted and contrast-reversed faces. Across five levels of external noise that spanned an order of magnitude, and which included the level of noise used in the experiments of this study, the noise-masking functions obtained with inverted and contrast-reversed faces were well fit by Equation 2.1 ($R^2 \geq 0.98$). Hence, our assumption that Equation 2.1 provides a good fit to thresholds in all conditions is reasonable.

For each level of stimulus contrast variance used during trials with external noise, a percentage of agreement, P_a, was calculated for replicated trials. A percentage of correct responses, P_c, was also estimated for each stimulus contrast by using the fitted Weibull (psychometric) function described earlier. By pairing P_a and P_c according to stimulus contrast, we were thus able to obtain a unique mapping between P_a and P_c. An observer modeled with different levels of internal noise, relative to a constant amount of externally added noise, responds with systematic changes to the slope s of this equation (Gold, Bennett & Sekuler, 1999b):

$$P_c = \log_{10}(P_a/100)s + 100 \tag{2.2}$$

The relationship between internal noise and s was measured by running Monte Carlo simulations of an observer performing in this experiment for 50 different levels of internal noise. By comparing a participant's slope to the modeled observer's slope, we were thus able to obtain an estimate of their total internal noise (σ_i), relative to external noise (σ_n). This noise standard deviation ratio

σ_i/σ_n was calculated for each participant, in all conditions.

2.2 Results

2.2.1 Contrast thresholds without external noise

The effects of stimulus inversion and contrast reversal were computed by dividing thresholds measured in the inverted, negative contrast, and combined (i.e, inverted and negative contrast) conditions by threshold in the upright, positive contrast condition (see Figure 2.5). A within-subject ANOVA on the log-transformed ratios revealed no significant effects (stimulus gender: $F(1,6) = 0.608$, $p = 0.46$; condition: $F(2,12) = 3.08$, $p = 0.083$; stimulus gender x condition: $F(2,12) = 1.21$, $p = 0.33$). An overall index of inversion and contrast reversal was estimated by averaging the three log-transformed ratios measured for each subject. The mean log-transformed ratio of 0.245 (i.e., mean ratio = 1.76) was significantly greater than zero ($t(6) = 6.27$, $p = 0.0007$). In other words, our observers demonstrated the expected impairment of performance when viewing inverted and contrast-reversed faces (for inverted faces, see Yin, 1969, Sekuler et al., 2004, and Martelli et al., 2005; for contrast-reversed faces, see Galper, 1970, Liu & Chaudhuri, 1997, Liu, Collin & Chaudhuri, 2000, Vuong et al., 2005, and Russell et al., 2006). Compared to normal faces, observers needed approximately 50% more contrast to identify inverted faces and 67% more contrast to identify contrast-reversed faces (see Figure 2.5). The magnitude of these effects fall within the range that Martelli, et al. (2005) report in their meta-analysis of face inversion effects. This result is important because it shows that our 2-alternative face recognition task elicits an inversion effect as strong as that obtained in a wide range of previously studied tasks (including tasks using a relatively large number of faces).

2.2.2 Noise-masking functions

To make it easier to compare the current findings with previous studies, equivalent input noise variance was converted to power spectral density by

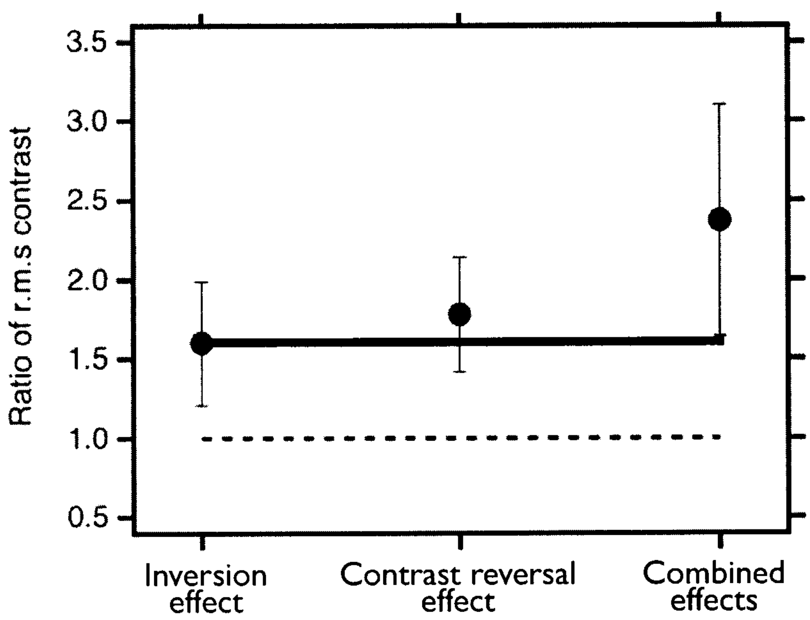


Figure 2.5: Ratios of rms contrast thresholds for recognition without display noise: inverted, contrast-reversed, and both inverted and contrast-reversed, each divided by normal-face thresholds. The data have been collapsed across stimulus gender. Error bars represent 95% confidence intervals. The dashed line indicates no effect of inversion or contrast reversal. The solid line indicates a contrast ratio of 1.6, which is the size of the average inversion effect reported in a meta-analysis by Martelli, Majaj and Pelli (2005).

multiplying σ_c^2 by the area of a single pixel (in deg^2). Equivalent input noise, averaged across all conditions, was 1.2×10^{-6} deg^2, which is very similar to the value reported in previous studies of 10-alternative face identification (Gold, Sekuler & Bennett, 2004; Gold, Bennett & Sekuler 1999b), 2-alternative face identification (Gold et al., 2004), and letter identification when the letters are the same size as the faces used in this study (Pelli & Farell, 1999). Equivalent input noise is plotted as a function of stimulus orientation and contrast polarity in Figure 2.6a. A 3-factor (Stimulus Gender x Orientation x Contrast Polarity) within-subjects ANOVA on log-transformed values of equivalent input noise found no significant main effects ($F < 1$, $p > 0.48$ in all cases) and no significant interactions. Hence, we found no evidence that equivalent input noise was affected by stimulus gender, orientation, or contrast polarity.

Log-transformed high-noise efficiency is plotted as a function of stimulus orientation and contrast polarity in Figure 2.6b. High-noise efficiency appeared to be significantly greater for upright than inverted faces, and for normal contrast than reversed-contrast faces. These observations were confirmed by a 3-factor (Stimulus Gender x Orientation x Contrast Polarity) within-subjects ANOVA on log-transformed high-noise efficiency, which revealed significant main effects of Orientation ($F(1,6) = 77.62$, $p = 0.00012$) and Contrast Polarity ($F(16) = 23.46$, $p = 0.0028$), and a significant Orientation x Contrast Polarity interaction ($F(1,6) = 7.91$, $p = 0.031$). No other effects were significant. After collapsing high-noise efficiencies across stimulus gender, multiple, one-tailed t-tests were performed to analyze the Orientation x Contrast Polarity interaction. Efficiency was higher for normal faces than for i) inverted faces ($t(6) = 7.64$, $p < 0.001$, one-tailed); ii) contrast-reversed faces ($t(6) = 4.82$, $p = 0.0014$, one-tailed); and iii) faces that were both contrast-reversed and inverted ($t(6) = 7.38$, $p < 0.001$, one-tailed). In addition, efficiency for inverted, positive contrast faces was higher than for inverted, negative contrast faces ($t(6) = 3.6$, $p = 0.006$, one-tailed). No other pairwise comparisons were significant. The significant Orientation x Contrast Polarity interaction implies that the effects of the two manipulations on high-noise efficiency were not additive.

Indeed, inspection of Figure 2.6b shows that the effects were sub-additive: inverting a contrast-reversed face reduced efficiency, but by an amount that was less than the prediction generated by adding the main effects of contrast polarity and orientation (i.e., the hollow triangle in Figure 2.6b).

2.2.3 Multiplicative noise

The relation between response accuracy and response consistency was similar across conditions (Figure 2.7a). Using the procedure outlined in the Methods section, internal:external noise ratios were estimated from the double-pass consistency data of all our observers (see Figure 2.6c). A 3-factor (stimulus gender X orientation X polarity) within-subjects ANOVA on log-transformed data revealed no significant effects ($F < 2.33$, $p > 0.17$ in all cases). Therefore, we found no evidence that multiplicative noise changed reliably with either stimulus inversion or contrast-reversal.

Linear regression was used to determine if individual differences in the effects of orientation (or contrast polarity) on efficiency were related to the effects of orientation (or contrast polarity) on internal:external noise ratios. For both positive and negative contrasts, the effect of orientation on efficiency — defined as the log difference between efficiencies measured in the upright and inverted conditions — were not related to the log difference between internal:external noise ratios measured in the upright and inverted conditions (positive contrast: $F(1,\ 12) = 0.075$, $p = 0.789$; negative contrast: $F(1,\ 12) = 2.128$, $p = 0.17$). Likewise, for both upright and inverted faces the effect of contrast polarity on efficiency was unrelated to the effect of contrast polarity on internal:external noise ratio (upright: $F(1,\ 12) = 1.728$, $p = 0.213$; inverted: $F(1,\ 12) = 0.382$, $p = 0.548$). Hence, we found no evidence that the effects of orientation and contrast polarity on high-noise efficiency were correlated with changes in internal:external noise ratio.

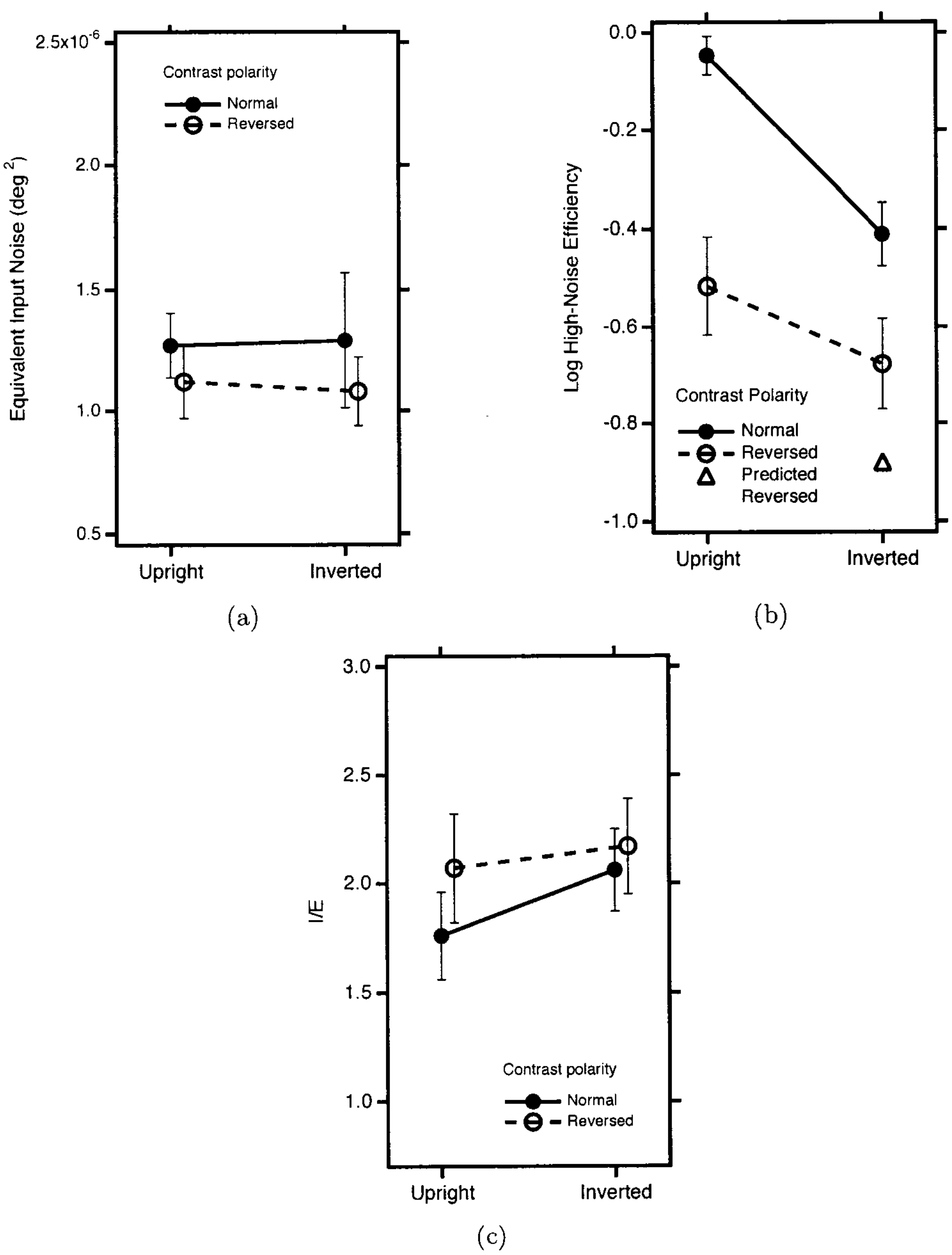


Figure 2.6: The effect of inversion and contrast-reversal on: a) equivalent input noise; b), log high-noise efficiency, and c), the internal-to-external noise ratio (I/E). Data averaged across observer and stimulus gender. Error bars represent ±1 SEM.

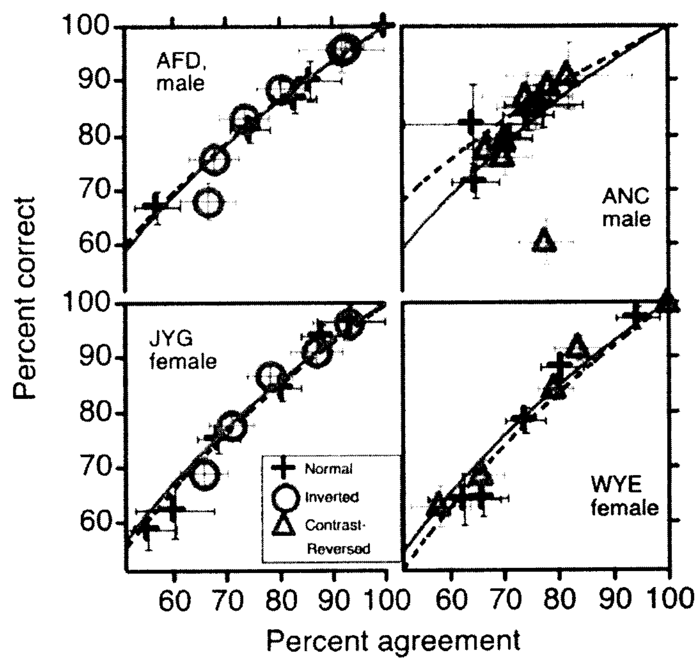


Figure 2.7: Top: Response consistency data for four observers. Each plot compares consistency measured in the normal (upright, positive-contrast) stimulus condition and either the inverted (left column) or contrast-reversed (right column) condition. Inversion had no significant effect on consistency.

2.3 Discussion

Face inversion and contrast reversal were associated with significantly higher identification thresholds and reduced high-noise efficiency. However, we found no evidence that inversion or contrast reversal affected the levels of contrast-independent and contrast-dependent internal noise. These results suggests that, compared to normal face recognition, observers are less able to use the available information when recognizing inverted or contrast-reversed faces.

Sekuler et al. (2004) demonstrated that observers used information conveyed by pixels near the eyes and eyebrows to identify upright and inverted faces. Nonetheless, there were subtle differences in the way these pixels were combined, and these differences were strongly correlated with the effect of face orientation on performance. Sekuler et al. argued that the subtle differences in classification images reflected differences in the way observers sampled information in upright and inverted faces. An alternative hypothesis, however, is that the structural differences in the classification images were caused by differences in internal noise (Murray et al., 2002). The results of the present study rule out this alternative explanation. Instead, the current findings are consistent with the idea that face inversion causes observers to make subtle, but consistent, changes in how they extract information from around the eye(s) to identify a face.

The failure to find a change in internal noise implies that rotating a face, or inverting its contrast, produces *systematic*, rather than stochastic, changes in face processing.In this regard, it is interesting to note that the differences between identification of normal and inverted (and/or contrast-reversed) faces are the same as those found between the identification of complex patterns after and before extended practice. Gold et al. (1999b; 2004) showed that practice in face and texture identification tasks increased high-noise efficiency but did not affect internal noise. Therefore, the specific kind of advantage we possess for upright face recognition, relative to inverted or contrast-reversed faces, is the same advantage that is developed with extensive practice. Additionally, Gold et al. (2004) measured classification images for 2-IFC face identification both before and after extensive practice. A comparison of pre- and post-learning classification images revealed that subtle rather than gross changes in the spatial sampling strategy were produced by perceptual learning. In direct parallel to the classification images for upright and inverted faces measured by Sekuler et al. (2004), observers in the Gold et al. (2004) study always appeared to be relying on pixels around the eyes and eyebrows regardless of how long they had been practicing. Faces that are both contrast-reversed and inverted are rarely, if ever, encountered in the natural environment. The results of our study, combined with those of Gold et al. (1999b, 2004), suggest that our experience with faces results in a more efficient information-sampling strategy, but only for faces presented in the familiar orientation and contrast polarity.

Rotating a face has a variety of effects on face perception: it makes it more difficult to detect misoriented facial features (Lewis, 2001; Thompson, 1980), to discriminate differences in the distances between facial features (Barton, Keenan & Bass, 2001; Freire, Lee & Symons, 2000; Leder & Bruce, 2000; Leder et al., 2001), as well as making it more difficult to identify the face. It is tempting to conclude that these various inversion effects are different manifestations of a single, underlying effect of orientation on face perception. However, it is possible that face inversion reduces the accuracy of identification and these other types of judgments for different reasons. The current study demonstrates that the sole effect of face inversion and contrast reversal on face identification is to reduce the sampling efficiency. It is not known whether face inversion, or contrast reversal, reduces the accuracy of other facial judgments for the same reason, or whether inversion and/or contrast reversal increases internal noise. To properly compare inversion and contrast reversal effects across stimuli and tasks, future studies should disentangle the effects of those manipulations on efficiency and internal noise.

2.3.1 Conclusion

The current experiments demonstrate that face inversion and contrast reversal cause a reduction in high-noise efficiency but have no effect on internal noise. These results have two implications for our understanding of face identification. First, they are consistent with the claim that inversion reduces face identification accuracy by producing a subtle but consistent shift observers' spatial-sampling strategies (Sekuler et al., 2004). Second, they implicate the role of orientation- and contrast-specific expertise in the face inversion and contrast-reversal effects.

2.4 References

Ahumada, A.J. (2002). Classification image weights and internal noise level estimation. Journal of Vision, 2, 121–131.

Barton, J.J., Keenan, J.P., & Bass, T. (2001). Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. British Journal of Psychology, 92 Part 3, 527-549.

Barton, J.J., Radcliffe, N., Cherkasova, M.V., Edelman, J., & Intriligator, J.M. (2006). Information processing during face recognition: The effects of familiarity, inversion, and morphing on scanning fixations. Perception, 35, 1089-1105.

Bennett, P.J., Sekuler, A.B., & Ozin, L. (1999). Effects of aging on calculation efficiency and equivalent noise. Journal of the Optical Society of America A: Optics, Image Science,and Vision, 16(3), 654–68.

Brainard, D.H. (1997). The psychophysics toolbox. Spatial Vision, 10(4), 433–6.

Burgess, A. (1985). Visual signal detection. III. On bayesian use of prior knowledge and cross correlation. Journal of the Optical Society of America A: Optics and ImageScience,2(9), 1498–507.

Burgess, A.E. & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. Journal of the Optical Society of America A: Optics and Image Science, 5(4), 617–27.

Dosher, B.A., & Lu, Z.L. (2000). Mechanisms of perceptual attention in precuing of location. Vision Research, 40(10–12), 1269–1292.

Freire, A., Lee, K., & Symons, L.A. (2000). The face-inversion effect as a deficit in the encoding of configural information: direct evidence. Perception, 29 (2), 159-170.

Galper, R.E. (1970). Recognition of faces in photographic negative. Psychonomic Science,19, 207–208.

Gold, J.M., Bennett, P.J., & Sekuler, A.B. (1999a). Identification of band-pass filtered letters and faces by human and ideal observers. Vision Research, 39(21), 3537–60.

Gold, J.M., Bennett, P.J., & Sekuler, A.B. (1999b). Signal but not noise changes with perceptual learning. Nature, 402(6758), 176–8.

Gold, J.M., Sekuler, A.B., & Bennett, P.J. (2004). Characterizing perceptual learning with external noise. Cognitive Science, 28(2), 167–207.

Green, D.M. (1964). Consistency of auditory detection judgments. Psychological Review, 71, 392–407.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53 (2), 513-536.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. Perception, 30 (1), 73-83.

Legge, G., Kersten, D., & Burgess, A.E. (1987). Contrast discrimination in noise. Journal of the Optical Society of America A, 4(2), 391–406.

Liu, C.H., & Chaudhuri, A. (1997). Face recognition with multi-tone and two-tone photographic negatives. Perception, 26(10), 1289-1296.

Liu, C.H., Collin, C.A., & Chaudhuri, A. (2000). Does face recognition rely on encoding of 3-d surface? Examining the role of shape-from-shading and shape-from-stereo. Perception, 29(6), 729-743.

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

Lu, Z.L., Liu, C.Q., & Dosher, B.A. (2000). Attention mechanisms for multi-location first- and second-order motion perception. Vision Research, 40(2), 173–186.

Martelli, M., Majaj, N.J., & Pelli, D.G. (2005). Are faces processed like words? a diagnostic test for recognition by parts. Journal of Vision, 5(1), 58–70.

Maurer, D., Le Grand, R., & Mondloch, C.J. (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6(6), 255–260.

Murray, R.F., Bennett, P.J., & Sekuler, A.B. (2005). Classification images predict absolute efficiency. Journal of Vision, 5(2), 139–49.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. Vision Research, 39(23), 3824–33.

Pelli, D.G. (1981). Effects of visual noise. Unpublished Ph.D. thesis, University of Cambridge, Cambridge.

Pelli, D.G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), Vision: Coding and Efficiency (pp. 3–24). Cambridge, UK: Cambridge University Press.

Pelli, D.G., & Farell, B. (1999). Why use noise? Journal of the Optical Society of America A: Optics Image Science Vision, 16(3), 647–653.

Pelli, D.G. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision, 10(4), 437–42.

Raghavan, M. (1995). Sources of visual noise. Unpublished doctoral dissertation, Syracuse University, Syracuse.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. Perception, 35(6), 749-759.

Sekuler, A.B., Gaspar, C.M., Gold, J.M., & Bennett, P.J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. Current Biology, 14(5), 391–6.

Steinman, R.M., Cunitz, R.J., Timberlake, G.T., & Herman, M. (1967). Voluntary control of microsaccades during maintained monocular fixation. Science, 155(769), 1577–9.

Thompson, P. (1980). Margaret Thatcher: a new illusion. Perception, 9 (4), 483-484.

Tjan, B.S., Braje, W.L., Legge, G.E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. Vision Research, 35(21), 3053–3069.

Tyler, C.W., Chan, H., Liu, L., McBride, B., & Kontsevich, L. (1992). Bit-stealing: How to get 1786 or more gray levels from an 8-bit color monitor. In: B.E. Ragowitz (Ed.) Human vision, visual processing, and digital display III. Proceedings of SPIE, 1666 (pp. 351-364).

Tyler, C.W. & Chen, C.C. (2000). Signal detection theory in the 2afc paradigm: attention, channel uncertainty and probability summation. Vision Research, 40(22), 3121–44.

Vuong, Q.C., Peissig, J.J., Harrison, M.C., & Tarr, M.J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and greebles. Vision Research, 45(10), 1213–23.

Wilson, H.R. (1991). Model of peripheral and amblyopic hyperacuity. Vision Research, 31(6), 967–82.

Yin, R.K. (1969). Looking at upside-down faces. Journal of Experimental Psychology, 81(1), 141-145.

Chapter 3

Spatial Frequency Tuning of Upright and Inverted Face Identification

Aside from the written word, there is no class of objects that the adult human has greater experience identifying than the human face. Of necessity, we become expert face identifiers and can recognize thousands of faces at a single glance (Bahrick, Bahrick & Wittlinger, 1975). However, if those same faces are turned upside-down, recognition becomes much more difficult (Yin, 1969; for a review, see Valentine, 1988). This result, known as the inversion effect, is interesting because an inverted face contains the same amount of information as an upright face. Thus, inverted faces must be more difficult to identify because the way we use that information varies as a function of orientation.

Previous results suggest that observers are less efficient at extracting the relevant information from inverted faces compared to upright faces (Sekuler, Gaspar, Gold & Bennett, 2004; Sekuler, Gold, Gaspar & Bennett, 2001), but the orientation-related changes in processing are subtle. For example, using the response classification technique, Sekuler et al. (2004) showed that observers used similar, localized regions of the face (i.e., near the eyes and eyebrows) to discriminate pairs of faces, regardless of face orientation. There also was no evidence for a difference in the extent of second order processing across face orientation. Hence, the difference in performance measured with upright and inverted faces reflected the fact observers used information around the eyes more optimally when the faces were upright, and not that different parts of the stimulus were used at different orientations.

In this paper we examine whether larger effects of orientation would be found in the spatial-frequency domain rather than the space domain. Faces are broadband patterns, containing information at all spatial scales (Gold, Bennett & Sekuler, 1999; Näsänen, 1999). Nonetheless, upright face identification appears to rely most heavily on a narrow band of spatial frequencies of approximately 6-13 cycles per face (Costen, Parker & Craw, 1996; Gold, et al., 1999; Näsänen, 1999; Ojanpää & Näsänen, 2003). It is not known, however, if a similar narrow band of frequencies mediates inverted face perception.

Previous researchers have suggested that processing of upright and inverted faces is qualitatively different — with configural processes dominating upright face recognition and feature-based processing dominating inverted face recognition (Leder, Candrian, Huber, & Bruce, 2001; Maurer, Le Grand, & Mondloch, 2002; Schwaninger, Lobmaier & Collishaw, 2002). Researchers also have suggested that the two types of processing are mediated by different spatial frequency bands (Costen, Parker & Craw, 1994; Glass, Bradshaw, Day & Umilta, 1985; Goffaux, Hault, Michel, Vuong & Rossion, 2005; Schwaninger, et al., 2002; Sergent, 1984; White & Li, 2006), with configural and feature-based processing mediated, respectively, by low and high spatial frequency mechanisms. Given this sort of reasoning, one should expect to see clear differences in the frequency bands underlying upright and inverted face processing, with a shift in the relevant frequencies across the two conditions. The present study examined this issue by using critical band masking (e.g., Patterson, 1974) to estimate the spatial frequency channel used to identify upright and inverted faces.

3.1 Methods

3.1.1 Observers

Six observers (4 male, 2 female; average age = 20) participated in the experiment. Four were experienced psychophysical observers (AML, JLT, KNT and TMC), but all were naïve about purpose of the experiment. All observers had an uncorrected or corrected binocular Snellen Acuity of 20/20 or better.

3.1.2 Stimuli & Apparatus

Stimuli were generated by a Power Mac G4 computer, and presented on a Sony Trinitron GDM-F520 monitor using MATLAB and the Psychophysics Toolbox (Brainard 1997; Pelli 1997). Face stimuli were based on digitized photographs of 10 faces (5 male and 5 female) cropped to an oval window, excluding areas showing the chin and hair, including the hairline (see Gold, Bennett, and Sekuler (1999) for details). From the viewing distance of 100 cm, the height and width of each face subtended 3.3 deg (96 pixels) and 2.3 deg (67 pixels), respectively. Faces were centered within a 4.4 x 4.4 deg square (128 x 128 pixels), and the amplitude spectrum of each individual face image was replaced with the average spectrum across all 10 images. Faces were presented to observers on a uniform background of average grey luminance or embedded within filtered or unfiltered white Gaussian noise (see Figure 3.1). Filtered noise was either low- or high-pass, each with cut-off frequencies of 1, 2.1, 4.2, 8.4 and 16.8 cycles per face width (cpf). The contrast variance of the unfiltered noise was 0.08.

3.1.3 Procedure

The sequence of events on a given trial was as follows: A small fixation square (8.3 x 8.3 arc min) appeared at the center of the screen (the polarity of

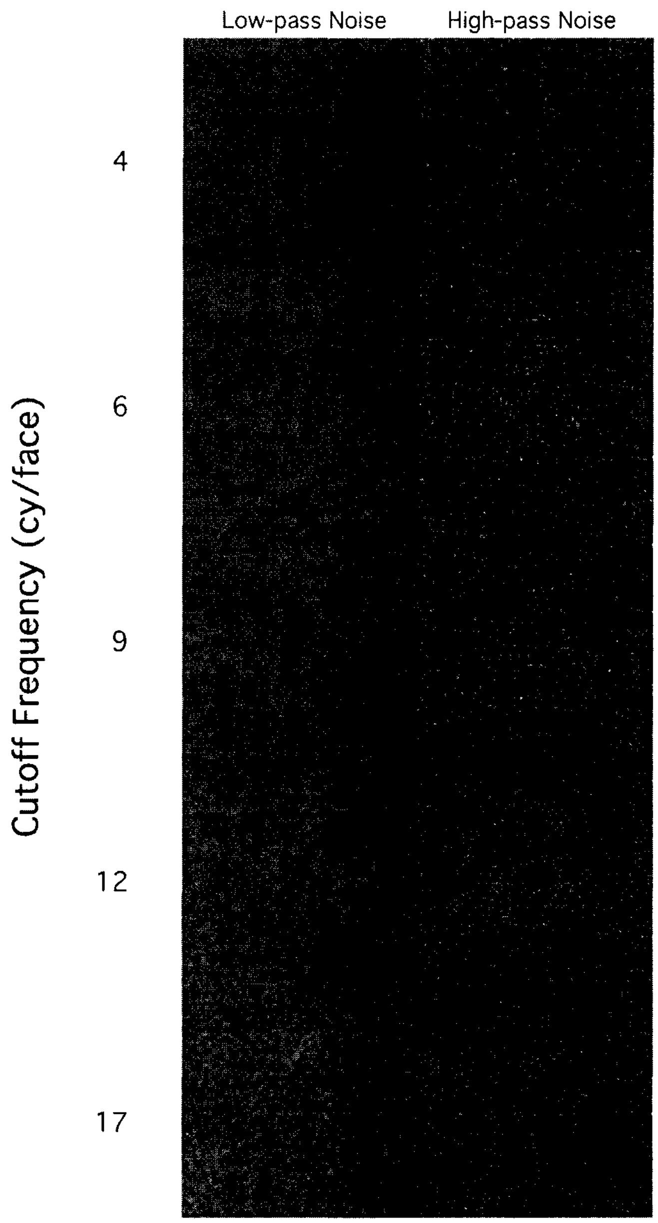


Figure 3.1: An illustration of the various types of noise used in the experiment. Two faces are embedded in low-pass and high-pass filtered noise, with cutoff frequency varying from approximately 4 to 17 cycles per face width (cpf). Face contrast variance is the same in every image. With low-pass noise, the face becomes difficult to identify when the cutoff frequency is above 9 cpf; with high-pass noise, the face is difficult to identify when the cutoff is below 9 cpf. In the experiment, cutoff frequencies varied from 1 to 32 cpf. In addition, faces were presented in all-pass (i.e., white) noise and in no noise (i.e., a uniform background).

the square - black or white - varied randomly from trial to trial). After 100 ms, the fixation square disappeared and a face was displayed for 500 msec. After the target face disappeared, observers were presented with a selection window that contained all ten faces at a r.m.s contrast of 0.32; each face was resized to 2.6 by 1.8 deg. The observer identified the target face by using the computer mouse to click on an item in the selection window. Auditory feedback, in the form of 600 and 200 Hz tones, was presented after correct and incorrect responses, respectively. After the response, the selection window disappeared, and the fixation square reappeared signaling the start of the next trial.

Observers participated in 3-6 one-hour sessions held on different days. Each session consisted of 12 or 14 blocks of trials, with a different noise presented in each block. All sessions contained the following stimulus conditions: no-noise, all-pass (white) noise, and low- and high-pass noise with cutoff frequencies of 1, 2.1, 4.2, 8.4 and 16.8 cpf. Some sessions included two additional blocks of all-pass noise. Also, thresholds were measured in two subjects with cutoff frequencies of 32 cpf. The order of conditions was randomized within a session. In each block, contrast variance of the face stimuli varied according to a 2-down/1-up staircase procedure, with an initial contrast variance of 0.001 and a constant step size of 0.25 log units; each staircase was terminated after 50 trials. Contrast thresholds for 71% accuracy were estimated for each condition by fitting a psychometric (Weibull) function to the results from each staircase.

Contrast variance was defined as $c^2 = \langle c^2(x, y) \rangle_{x,y}$, where $\langle \; \rangle$ indicates expected value (over all x, y), $c(x, y) = (L(x, y) - L_{background}) / L_{background}$, and $L(x, y)$ is the luminance at location (x, y). Final estimates of contrast thresholds were calculated by averaging thresholds obtained in the last three sessions. These average thresholds were then used to derive a set of four noise-masking functions for each of the six observers: upright face recognition in low- and high-pass noise, and inverted face recognition in low- and high-pass noise.

3.1.4 Data Analysis Overview

Four spatial-frequency channels were estimated for each of the 6 observers: channels for upright face recognition in low- and high-pass noise, and inverted face recognition in low- and high-pass noise. Assuming that the channel is linear, the integral of the channel's tuning function can be used to predict the shape of the measured noise masking functions (see Patterson (1974) for details). Parameters describing the spatial frequency tuning function for each channel were chosen such that the resulting predictions of contrast threshold best fit the data. The analysis of the best-fitting channels enabled us to compare tuning parameters across orientation and noise type. Bootstrap analyses were performed to obtain robust estimates of channel parameters, and their confidence intervals, for each individual observer. The analysis of the confidence intervals enabled us to gauge the effect of stimulus orientation and noise type for each observer.

3.1.5 Spatial-frequency Channels

We assumed that observers based their decisions on visual information that has passed through a linear spatial frequency filter. Spatial frequency tuning was modeled using a modified lognormal probability density function:

$$G(f) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\log_{10}(f/\mu)^2/2\sigma^2} \tag{3.1}$$

where f refers to spatial frequency, G is the gain of the channel, and σ and μ are free parameters determining, respectively, the bandwidth and peak sensitivity of the channel. Peak sensitivity refers to the spatial frequency at maximum channel gain.

3.1.6 Noise-masking Functions

Thresholds from each condition were used to derive four noise-masking functions: thresholds for upright face identification in low- and high-pass noise, and inverted face identification in low- and high-pass noise. Noise-masking functions were based on the integral of G in Equation 3.1. We assumed that spatial frequencies less than 0.1 cpf, and greater than 100 cpf, did not contribute significantly to performance, and therefore used those frequencies as the limits of integration. The following equation was used to predict contrast thresholds measured in low-pass noise with a cut-off at f cpf:

$$c_{pred}^2(f) = \int_{0.1}^{f} G(f) \times (c_{all}^2 - c_{no}^2) + c_{no}^2 df \tag{3.2}$$

where is predicted contrast threshold, and and are thresholds measured in all-pass noise and in no noise, respectively. The equation used for high-pass noise is the same, except that the integrals were performed from 100 cpf down to a cut-off of f cpf:

$$c_{pred}^2(f) = \int_{f}^{100} G(f) \times (c_{all}^2 - c_{no}^2) + c_{no}^2 df \tag{3.3}$$

3.1.7 Fitting Channel to Noise-masking Function

The MATLAB optimization toolbox was used to choose σ and μ such that they minimized the following measure of error:

$$\text{error} = \sum_f \log_{10} \left(\frac{c(f)^2}{c(f)_{pred}^2} \right)^2 \tag{3.4}$$

As described earlier, f took on the values 1, 2.1, 4.2, 8.4 and 16.8 cpf. To constrain c_{pred}^2 to approach and c_{all}^2 at reasonable cut-off frequencies, we treated c_{all}^2 and c_{no}^2 as thresholds measured at 0.1 and 100 cpf, respectively. Therefore, each noise-masking function was composed of 7 thresholds. We constrained the fits so that the bandwidth parameter, σ, was between 0.01 and 2, and peak frequency, f, to be between 0.1 and 30 cpf.

3.1.8 Bootstrap Analysis

Resampling methods were used to calculate confidence intervals for σ and μ (Efron & Tibshirani, 1998). Deviations from the derived noise-masking functions were assumed to reflect random sampling noise, or error. For each threshold, the relative error was defined as:

$$\text{err} = \left(c(f)^2 - c(f)^2_{pred}\right) / c(f)^2_{pred} \tag{3.5}$$

The magnitude of the relative error did not differ in any obvious way across conditions. By randomly sampling, with replacement, from this set of errors, we created bootstrapped sets of thresholds using the equation:

$$c(f)^2_{boot} = \text{err}(i) \times c(f)^2_{pred} + c(f)^2_{pred} \tag{3.6}$$

where i is a random index into $\text{err}(f)$. Equations 3.2 and 3.3 were then used to derive bootstrapped noise masking functions that fit the bootstrapped thresholds.

For each predicted noise-masking function, a family of 1000 bootstrapped noise-masking functions were generated, each generating in turn a bootstrapped value of σ and μ. In this way, we obtained a distribution of 1000 values for both parameters in each condition. The means and 95-percent confidence intervals for σ and μ were calculated from these distributions.

3.2 Results

3.2.1 Classic Face Inversion Effects

The magnitude of the face inversion effect was estimated by dividing threshold in the inverted-face condition by threshold in the upright-face condition: the average threshold ratios were 1.73 (se = 0.24) and 1.83 (se = 0.21) in the no noise and all-pass noise conditions, respectively. In both noise conditions, a one-tailed t-test on the log-transformed threshold ratios rejected the null hypothesis of no difference between face orientations (no noise: $t = 3.99$, $df = 5$, $p < .01$; all-pass noise: $t = 4.67$, $df = 5$, $p < .01$). Moreover, the magnitude of the face inversion effect was nearly equal to the inversion effect obtained by Sekuler et al. (2004), and to the average value calculated by Martelli et al. (2005) in a meta-analysis of 16 face recognition studies. Hence, the observers in the current experiment exhibited inversion effects that were comparable to those reported in previous face perception studies.

Figure 3.2 shows the average inversion effect obtained in each condition. The leftmost and rightmost points in Figure 3.2 depict inversion effects obtained with no noise and all-pass (i.e., white) noise, respectively; the remaining symbols indicate inversion effects obtained with low-pass and high-pass noise at various cutoff frequencies. Only two subjects were tested with a cutoff frequency of 32 cpf, which partly accounts for the larger standard errors in those conditions. Data from all of the other conditions were analyzed with a repeated-measures ANOVA: The effect of noise condition was not significant ($F(11, 55) = 0.502$, $MSE = 0.012$, $p = 0.89$), demonstrating that the size of the inversion effect did not vary significantly across conditions. This finding implies that the mechanisms that produce the inversion effect were affected similarly by manipulations of the signal-to-noise ratio in various bands of spatial frequencies, and is consistent with previous reports that the magnitude of the inversion effect is not affected by band-pass filtering faces at different center frequencies (Boutet et al., 2003). Moreover, the results in Figure 3.2 imply that the shapes of the masking curves were similar in the upright and inverted conditions, and therefore that the channels derived from those functions should be similar. We examine this issue quantitatively in the next section.

3.2.2 Noise-masking & Tuning Functions

Figure 3.3 depicts noise-masking and tuning functions from one observer whose results are representative of those obtained in the experiment. Noise-masking functions are shown in the top rows and tuning functions in the bottom rows; left and right columns depict results for upright and inverted

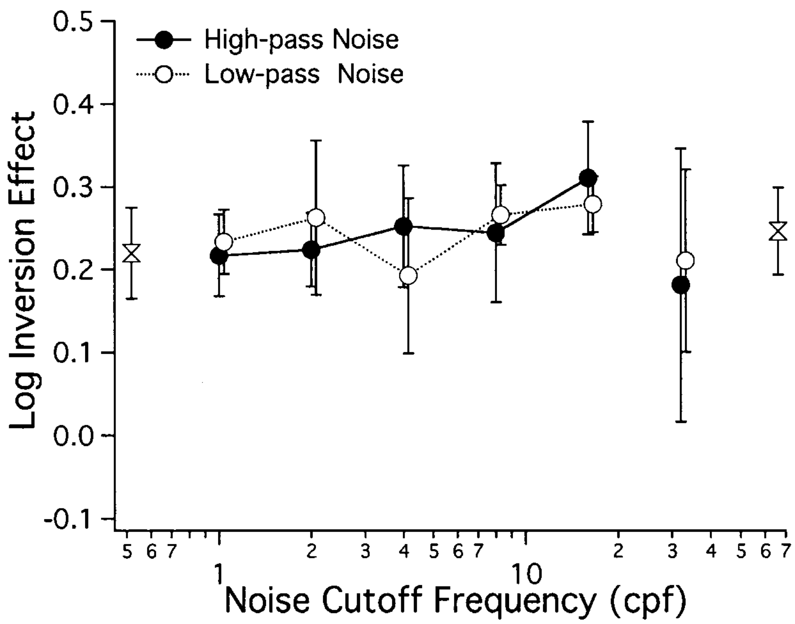


Figure 3.2: Average inversion effects plotted as a function of the cutoff frequency of the noise. The inversion effect was defined as the logarithm of ratio of thresholds, expressed as contrast, obtained with upright and inverted faces. The leftmost and rightmost symbols represent inversion effects measured in the no-noise and all-pass noise conditions. The points at 32 cpf represent the mean of the two subjects tested in that condition; the remaining points represent the mean of six subjects. Error bars represent ±1 SEM. Inversion effects did not differ significantly across conditions.

faces, respectively. To make the visual comparison between the shapes of low- and high-pass tuning functions easier, the tuning functions have been normalized to a maximum value of one. In conditions of low-pass noise, contrast thresholds rose as a sigmoidal function of cut-off frequency; in high-pass noise, thresholds decreased as a sigmoidal function of cut-off frequency. Across all conditions and observers, the noise masking functions fit the threshold data reasonably well (see Table 3.1). As described above, the spatial frequency tuning functions are the derivatives of the noise masking functions. Boxplots of tuning parameters are shown in Figure 3.4: although the peak frequencies and bandwidths (i.e., full width at half-height) of the tuning functions varied significantly across observers, the median tuning parameters (i.e., peak frequency = 9 cpf; bandwidth = 2 octaves) are similar to previous estimates obtained for upright faces using different psychophysical methods (Gold, Bennett & Sekuler, 1998; Gold et al., 1999; Näsänen, 1999; Ojanpää & Näsänen, 2003).

	Upright Faces		Inverted Faces	
Observer	High-pass	Low-pass	High-pass	Low-pass
AML	99.38	97.66	97.72	99.26
AND	97.98	96.27	97.20	95.17
JNG	83.16	97.52	92.40	97.29
JUL	99.63	98.12	99.43	99.62
KNM	97.34	98.72	80.71	96.91
TMC	98.74	96.16	96.73	98.38

Table 3.1: Percentage R^2 between measured and predicted noise-masking functions. All of the nose-masking were well fit by cumulative lognormal density functions..

3.2.3 Confidence intervals of channel parameters

Confidence intervals for peak tuning frequency and bandwidth are shown in Figure 3.5. For every observer, the confidence intervals for both parameters obtained with upright and inverted faces overlapped. Moreover, neither peak frequency nor bandwidth appeared to vary consistently with stimulus orienta-

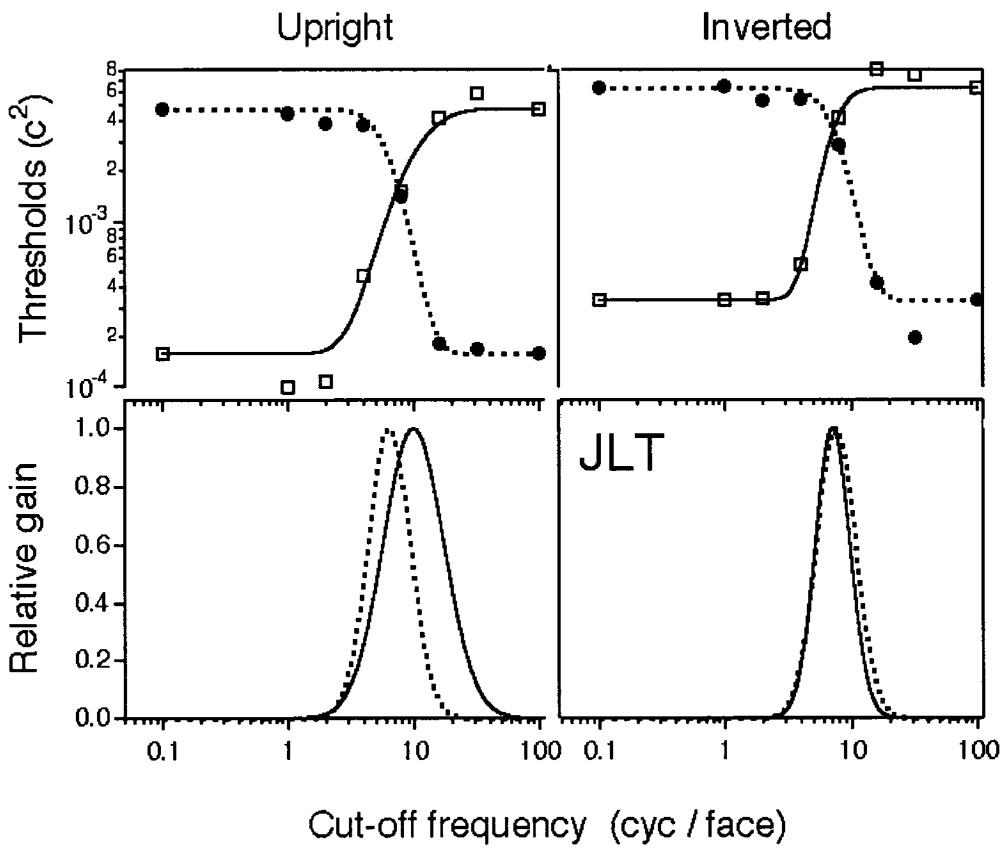
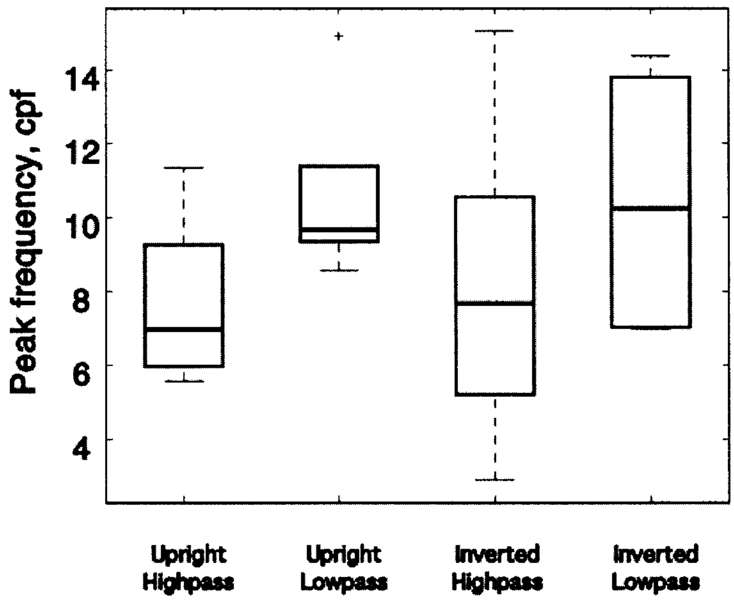
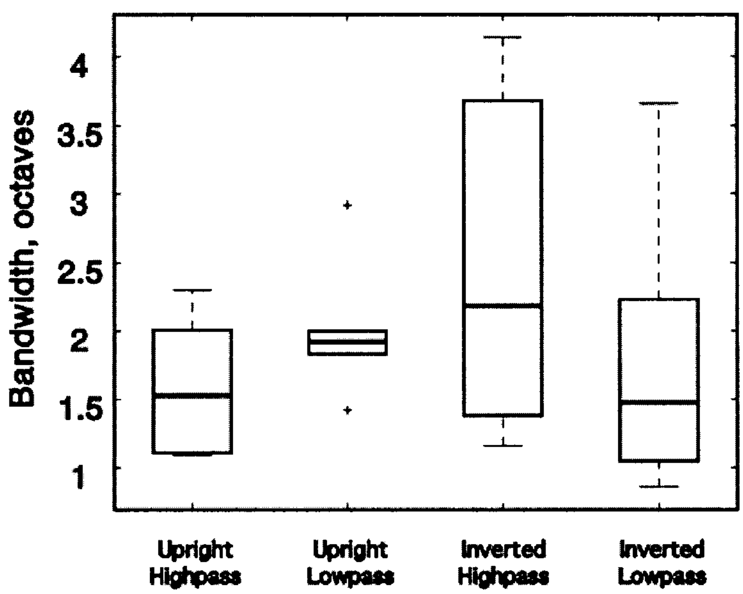


Figure 3.3: Results from one observer that are representative of those obtained from all observers with upright and inverted faces (left and right columns, respectively). The upper panels show thresholds obtained with faces embedded in high-pass (filled circles) and low-pass (unfilled squares) noise, and the best-fitting noise-masking functions. The spatial frequency channels, shown in the lower panels, were obtained by taking the derivatives of the corresponding noise-masking functions in the upper panels (dashed lines for high-pass, and solid lines for low-pass noise). The peak of the channels have been normalized to one. The data suggest that this observer used a narrow-band spatial frequency channel ($\approx$ 1.3 octaves) to identify both upright and inverted faces. Peak spatial frequency is very similar across all four conditions (7.6 cpf on average), but differs slightly between noise conditions for upright faces. However, the effect of noise type on peak frequency was not statistically reliable across observers.



(a)



(b)

Figure 3.4: Boxplots of peak spatial frequency (a) and bandwidth (b) across observers, in each of the 4 stimulus conditions. Bandwidth is the full-width at half-height, expressed in octaves.

tion. These observations were evaluated quantitatively in the following way. For each observer, the estimated tuning parameters at each cutoff frequency were averaged across low- and high-pass noise conditions, and the difference between the average in the upright and inverted conditions was calculated. For each tuning parameter, t-tests were used to evaluate the null hypothesis that the mean difference score did not differ from zero. Neither t-test was significant (peak frequency: 95% confidence interval = (-1.9,1.6), t(5) = -0.173, p=0.87; bandwidth: 95% confidence interval = (-1.27,0.62), t(5) = -0.883, p=0.42). Essentially identical results were obtained if, instead of averaging across noise conditions, separate tests of face orientation were done on tuning parameters estimated in the low- and high-pass noise conditions. Finally, the same pattern of results was obtained when the analyses were repeated using a percentile-t bootstrap method that did not require the assumption that the difference scores were distributed normally. In summary, there was no evidence that either tuning parameter - peak frequency or bandwidth - varied consistently with stimulus orientation.

Next, we examined whether differences in peak frequency or bandwidth were related to the size of the inversion effects measured with no noise or with all-pass (i.e., white) noise. Linear models were created that related the inversion effect (i.e., threshold in the inverted condition divided by threshold in the upright condition) to the sum of three predictor variables: the difference between peak frequencies estimated with upright and inverted faces, the difference between bandwidths, and the interaction between the peak frequency and bandwidth predictor variables. None of the regression coefficients for the predictor variables differed significantly from zero ($\text{abs}(t) < 0.7$, $df = 1$, $p > 0.55$, in all cases). Essentially the same results were obtained when the linear model was fit to the log-transformed inversion effects. Hence, we found no evidence that the size of the inversion effect was related to differences between the tuning parameters estimated with upright and inverted faces.

3.2.4 Off-frequency looking

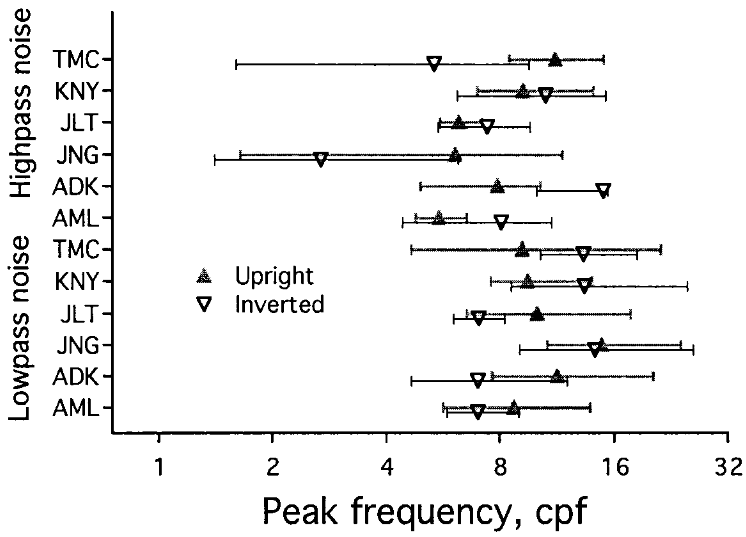
Thus far, we have discussed how estimates of spatial-frequency tuning are derived by embedding faces in low- or high-pass noise and measuring how contrast thresholds vary as a function of the cut-off frequency of the filtered noise. However, estimating tuning in this way cannot provide meaningful results if observers alter their tuning according to the cut-off frequency of the noise. Rather than being fixed, an observer's spatial frequency tuning might be redirected flexibly toward regions of the spectrum that are not corrupted by external noise, a type of behavior known as off-frequency looking (Pelli, 1981). If our observers looked off-frequency, then our estimates of their tuning function would be a misleading composite of multiple tuning functions. Fortunately, we can assess the degree to which an observer engaged in off-frequency looking by measuring the Log Additivity Ratio (LAR) (e.g., Solomon & Pelli, 1994; Majaj, Pelli, Kurshan and Palomares, 2002).

The energy of a visual stimulus, E, is its contrast variance multiplied by area. The LAR is defined as

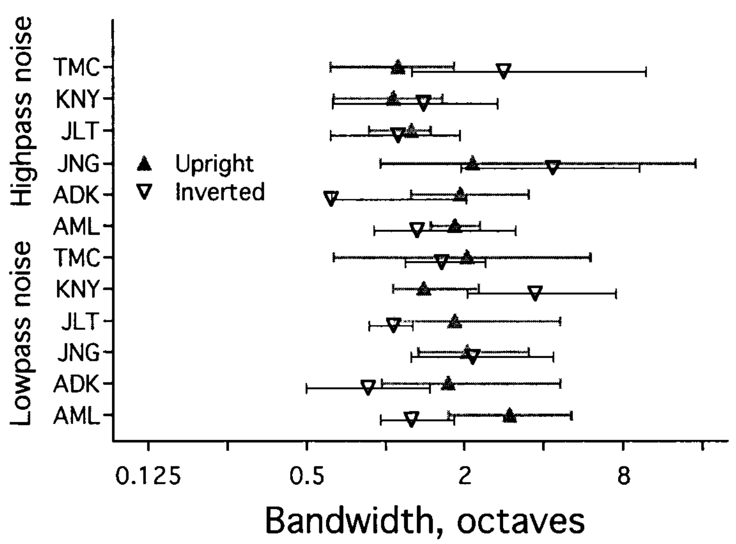
$$\mathrm{LAR} = \log\left(\frac{E^{+}_{low} + E^{+}_{high}}{E^{+}_{all}}\right)$$

where E^{+} represents threshold elevations. For example, E^{+}_{low} is the threshold energy for faces embedded in low-pass filtered noise, minus threshold energy for faces without noise. Similarly, E^{+}_{high} is the threshold energy for faces embedded in high-pass filtered noise, minus threshold energy for faces without noise. Finally, E^{+}_{all} is the threshold energy for faces embedded in all-pass noise, minus threshold energy for faces without noise. The cut-off frequencies for the low- and high-pass noises analyzed in a LAR are always the same, so that the range of noise included in $E^{+}_{low} + E^{+}_{high}$ is equivalent to the range of noise in E^{+}_{all}. If an observer uses a single, fixed linear spatial frequency channel when faces are embedded in low- and high-pass versions of noise with the same cut-off frequency, then the effects of all-pass noise on threshold should equal the sum of the effects of low-pass and high-pass noise, $E^+_{all} = E^+_{low} + E^+_{high}$, and therefore the value of LAR will be zero. However, if an observer can focus on spatial frequencies that are not embedded in noise — perhaps by using channels that are shifted toward higher frequencies when stimuli are embedded in low-pass noise, and toward lower frequencies when stimuli are in high-pass noise — then the sum of the effects of low- and high-pass noise should be less than the effects of all-pass noise, resulting in a negative LAR. The bottom-left panel of Figure 3.3 shows the spatial-frequency channels for observer JLT viewing upright faces. Notice how the peak spatial-frequency for faces embedded in low-pass noise (10 cpf) is slightly higher than for faces in high-pass noise (6 cpf). This difference in peak frequency is what would be expected if the observer adjusts frequency selectivity to improve the signal-to-noise ratio, and the value of LAR is -0.5 at a cut-off of 8 cpf.

Inspection of Figure 3.4a shows that, across all observers, the median peak frequency was slightly higher in the low-pass noise condition than in the high-pass noise condition. Although these trends were not statistically significant, they are suggestive of off-frequency looking. We therefore conducted an analysis of variance on LAR scores to assess the magnitude of off-frequency looking. For each observer, a LAR value was calculated at each cut-off frequency using thresholds taken from the functions fit to the low-pass and high-pass noise masking data. The ten LAR values (five at each orientation) were then submitted to a 2 (stimulus orientation) x 5 (cut-off frequency) repeated-measures ANOVA. None of the effects was significant (orientation: $F(1, 45) = 0.18$, $p = 0.68$; cut-off frequency: $F(4, 45) = 1.82$, $p = 0.14$; orientation x cut-off frequency: $F(4, 45) = 0.33$, $p = 0.86$), indicating that the value of LAR did not vary across conditions. Furthermore, the average LAR did not differ significantly from zero ($M = -0.127$, $F(1, 45) = 2.89$, $p = 0.10$), suggesting that observers did not engage in significant amounts of off-frequency looking.



(a)



(b)

Figure 3.5: Bootstrapped, 95% confidence intervals for peak frequency (a) and bandwidth (b) for each observer in each condition. The top and bottom parts of each panel show the results from the high-pass and low-pass noise conditions, respectively. Gray, upward-pointing triangles depict median values for upright faces; and black, downward-pointing triangles depict median values for inverted faces. All of the confidence intervals overlap, suggesting that stimulus orientation does not influence spatial frequency tuning in a reliable fashion.

3.3 Discussion

Observers use a narrow band of spatial frequencies to discriminate and identify upright faces (Costen et al., 1996; Gold et al., 1999; Näsänen, 1999; Ojanpää & Näsänen, 2003). Previous researchers have suggested that configural and feature-based processing, which are thought to draw on information contained in different bands of spatial frequencies, contribute differentially to the perception of upright and inverted faces (Costen et al., 1994; Glass et al., 1985; Goffaux et al., 2005; Schwaninger et al., 2002; Sergent, 1984; White & Li, 2006). We therefore hypothesized that face inversion effects might reflect a difference in the spatial frequencies chosen to recognize upright and inverted faces. The results of the current study are inconsistent with this hypothesis. The current experiment replicated previous results concerning the spatial-frequency selectivity of upright face recognition: on average, observers relied on a 2 octave wide band of spatial frequencies, centered on about 9 cpf. However, there was no evidence that the spatial-frequency tuning of upright and inverted face recognition differed. First, there was no reliable difference in the peak spatial frequency used to identify upright and inverted faces. Second, the bandwidth of tuning for inverted face recognition was not reliably different from that of upright face recognition. Finally, the size of the inversion effect was not related to the spatial frequency content of the noise, nor to differences between the channel parameters (i.e., peak tuning frequency and bandwidth) estimated for upright and inverted faces. In summary, our noise masking results suggest that observers use approximately the same narrow band of spatial-frequencies to identify faces, whether the faces are shown upright or inverted: they do not support the hypothesis that upright and inverted faces are processed by mechanisms that encode information carried by different spatial frequency bands (e.g., Goffaux et al., 2005). Instead, our results are consistent with previous reports that the size of the inversion effect is unaffected by variations in the spatial frequency content of upright and inverted faces (Boutet et al., 2003).

Although observers use the same band of frequencies when recognizing either upright or inverted faces, it is possible that information within the same narrow frequency band is used differently when processing faces in different orientations. For example, suppose observers discriminate upright and inverted faces on the basis of the response of a local, band-limited spatial filter placed near an eye or eye-brow. Within this framework, our results suggest that the filters used to encode upright and inverted faces have the same spatial frequency tuning. However, slight misplacement or inappropriate rotation of the local filter relative to the inverted face generally would result in a lower signal-to-noise ratio, and therefore could contribute to the face inversion effect without altering the critical band for face processing. The results of a recent classification image study by Sekuler et al. (2004) are consistent with this idea. Sekuler et al. (2004) measured classification images for upright and inverted face discrimination and found that there were no obvious differences in the general regions of the face were correlated with behaviour: observers discriminated upright and inverted faces on the basis of the spatial distribution of contrast near the eyes and eyebrows. However, observers were more efficient at extracting information from those regions when upright faces were presented. Taken together, the results of Sekuler et al. (2004) and those of the current study lead to the suggestion that the processes involved in recognizing upright and inverted faces may be more similar than current theories suggest.

3.4 References

Bahrick, H., Bahrick, O., & Wittlinger, R. (1975). Fifty years of memory for names and faces: A cross-sectional approach. Journal of Experimental Psychology: General, 104(1):54–75.

Boutet, I., Collin, C., & Faubert, J. (2003). Configural face encoding and spatial frequency information. Perception & Psychophysics, 65(7):1078–93.

Brainard, D. (1997). The psychophysics toolbox. Spatial Vision, 10(4):443–6.

Costen, N.P., Parker, D.M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. Perception, 23(2):129–46.

Costen, N.P., Parker, D.M. & Craw, I. (1996) Effects of high-pass and low-pass spatial filtering on face identification. Perception & Psychophysics, 58, 602-612.

Efron, B. & Tibshirani, R. (1998). An Introduction to the Bootstrap. Boca Raton, FL: Chapman & Hall/CRC.

Glass, C., Bradshaw, J.L., Day, R.H., & Umilta, C. (1985). Familiarity, spatial frequency and task determinants in processing laterally presented representations of faces. Cortex, 21(4):513–31.

Goffaux, V., Hault, B., Michel, C., Vuong, Q.C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. Perception, 34(1):77–86.

Gold, J., Bennett, P.J., &Sekuler, A.B. (1998, May). Identification efficiency for low- and high-pass filtered letters and faces. Poster presented at the annual meeting of the Association for Research in Vision and Ophthalmology, Fort Luaderdale, FL, USA. Investigative Ophthalmology and Visual Science (Supplement), 39.

Gold, J., Bennett, P.J., & Sekuler, A.B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. Vision Research, 39(21):3537–60.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. Perception, 30(1):73–83.

Majaj, N., Pelli, D., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. Vision Research, 42(9):1165–84.

Martelli, M., Majaj, N., & Pelli, D. (2005). Are faces processed like words? a diagnostic test for recognition by parts. Journal of Vision, 5(1):58–70.

Maurer, D., Le Grand, R., & Mondloch, C. (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6(6):255–260.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. Vision Research, 39(23):3824–33.

Ojanpää, H. & Näsänen, R. (2003). Utilisation of spatial frequency information in face search. Vision Research, 43(24):2505-2515.

Patterson, R. (1974). Auditory filter shape. Journal of the Acoustical Society of America, 55(4):802–9.

Pelli, D.G. (1981) Effects of visual noise. Ph.D. thesis. Cambridge University, Cambridge, England. http://vision.nyu.edu/docs/Pelli-Thesis1981.pdf.

Pelli, D.G. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision, 10(4):437–42.

Schwaninger, A., Lobmaier, J.S., & Collishaw, S.M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. Lecture Notes in Computer Science, 2525:643–650.

Sekuler, A.B., Gaspar, C.M., Gold, J.M., & Bennett, P.J. (2001). The efficiency of face recognition: effects of inversion and contrast reversal. Paper presented at the European Conference for Visual Perception, Kusadasi, Turkey. Perception (ECVP2001 Supplement).

Sekuler, A.B., Gaspar, C.M., Gold, J.M., & Bennett, P.J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. Current Biology, 14(5):391–6.

Solomon, J. & Pelli, D. (1994). The visual filter mediating letter identification. Nature, 369, 395-397.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. British Journal of Psychology, 75:221–42.

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. British Journal of Psychology, 79(4):471–91.

White, M. & Li, J. (2006). Matching faces and expressions in pixelated and blurred photos. American Journal of Psychology, 119(1):21–8.

Yin, R. (1969). Looking at upside-down faces. Journal of Experimental Psychology, 81(1):141-145.

Chapter 4

Discrimination of distance between facial features

4.1 Introduction

Many researchers (Diamond & Carey, 1986; Gauthier, Williams, Tarr & Tanaka, 1998; Kemp, McManus, & Pigott, 1990; Le Grand, Mondloch, Maurer, & Brent, 2001; 2003; Leder & Bruce, 2000; Leder, Candrian, Huber, & Bruce, 2001; Maurer, Le Grand, & Mondloch, 2002; Rhodes, 1988; Searcy & Bartlett, 1996) have claimed that human face recognition relies critically on our ability to accurately judge the spacing among facial features. Both Searcy and Bartlett (1996) and Le Grand et al. (2001; 2003) have speculated that face recognition is in fact supported by a mechanism that is specialized for processing variations in feature-spacing, in addition to a mechanism for processing features individually. The perceived importance of feature-spacing as a source of information about facial identity has also influenced the design of stimulus sets that are meant to probe perceptual expertise (Gauthier, Williams, Tarr & Tanaka, 1998) and the underlying neural representation of faces (Wilson, Loffler & Wilkinson, 2002). The synthetic face stimuli designed by Wilson et al., (2002) are especially striking in their emphasis on the spatial arrangement of features: individual faces in that stimulus class vary only in facial geometry (including feature-spacing), and possess the same set of generic facial features, including the eyes and eyebrows. Steyvers and Busey (2001) used variations in the local shape and texture and variations in feature-spacing as separate dimensions to model human judgments of the similarity between individual faces. Finally, the influence of feature-spacing can also be seen in the field of development, where the face recognition expertise of the adult observer has been attributed to an enhanced ability to discriminate feature-spacing (Mondloch, Le Grand & Maurer, 2002). In conclusion, many researchers claim that the spatial arrangement of features is important for face recognition, and this claim has had a considerable influence on theories about face perception.

Despite the theoretical importance attached to feature-spacing, surprisingly little is known about whether the discrimination of feature-spacing can support face recognition in natural contexts. Multiple studies have measured thresholds for the discrimination of faces that differ in the distance between a pair of features, usually inter-pupillary distance, or IPD (Haig, 1984; Kemp, McManus, & Pigott, 1990; Leder, Candrian, Huber, & Bruce, 2001; Barton, Keenan & Bass, 2001). Like previous researchers we measure thresholds for the discrimination of IPD. For the first time however, we compare these thresholds against the actual variation of IPD in a real population of faces (Farkas, 1981; Young, 1993). If IPD, by itself, constitutes a reliable cue for facial identity then observer thresholds should be low relative to the natural range of IPDs (depicted in Figure 4.2). In order to obtain a single measure of the utility of IPD, based on both human thresholds and facial statistics, we ask the following question: if we randomly sampled two faces from the population, what is the probability that the resulting difference in IPD can be discriminated at threshold? A high value of this probability – which we refer to as the Probability of Encountering a Discriminable Pair, or PED – implies that threshold is low relative to the variation in IPD, and therefore that IPD may be useful for discriminating among naturalistic faces.

With two exceptions, every study that has measured feature-spacing thresholds has either used a task that requires observers to discriminate faces differing *only* by feature-spacing (Haig, 1984; Kemp, McManus, & Pigott, 1990; Leder, Candrian, Huber, & Bruce, 2001; Barton, Keenan & Bass, 2001). In natural viewing conditions, however, most faces differ in terms of both facial geometry (i.e., feature-spacing) and other, non-geometric, characteristics such as hair or eye color, or the thickness of eyebrows (Steyvers & Busey, 2001). Moreover, faces viewed at different distances form retinal images of varying size, forcing observers to take into account the actual size of the object when judging inter-feature spacing. It is quite possible that these variations in both facial identity and size interfere with the accurate discrimination of feature spacing. Like most previous studies, the current study measures discrimination thresholds for faces that differ only in IPD (Experiment 1). However, this study also measures thresholds for the discrimination of IPD among faces of different individuals (Experiment 2), and among faces that differ in both size and identity (Experiment 3). If the additional variation in facial identity, or the combined variation in facial identity and size, make IPD discrimination more difficult, there should be corresponding increases in IPD threshold. Barton, Zhao and Keenan (2003) measured the accuracy and speed of face discrimination in a condition where faces differed only in eye position, and also in a condition where they differed in both eye position and eye brightness. Barton et al. found that accuracy and reaction time did not differ across conditions. This result is consistent with the idea that variations in facial identity do not interfere with the discrimination of feature-spacing. However, variations in facial identity are accompanied by a wide range of physical changes besides differences in eye brightness. It should also be noted that both Barton et al. and Malcolm, Leung and Barton (2004) measured performance in feature-spacing discrimination tasks that required the observer to compare faces varying in overall size. However, those two studies do not also measure performance in a control condition where overall face size was held constant. Therefore, it remains unclear whether or not feature-spacing discrimination is worse when one is required to compare faces that vary in size. In this regard, PED values measured in Experiment 3 are especially informative because they indicate the utility of IPD discrimination in the most realistic conditions tested in this study.

In the population of Caucasian faces, nose-to-mouth distance (NMD) has a greater coefficient of variation than IPD (Farkas, 1981; Young, 1993), and therefore, NMD provides a more useful cue for face identification than does IPD. If humans are sensitive to the relative informativeness of different feature-spacing cues, then one would expect NMD thresholds to be lower than IPD thresholds. Therefore, we also measured thresholds for the discrimination of NMD in Experiment 4. As in Experiment 3, observers in Experiment 4 discriminated faces that differed additionally in identity and size. In order to determine the relative importance of IPD and NMD for realistic face identification, PED values for both tasks are compared. It is necessary to use a measure like PED to compare the utility of various feature-spacing cues because one can come to very different (and misleading) conclusions by using either thresholds or facial statistics alone.

As mentioned earlier, human faces provide signals about identity other than those related to the variation in feature-spacing (Steyvers & Busey, 2001; Zhao, Chellappa, Phillips & Rosenfeld, 2003). In fact, many computer vision solutions to face recognition employ some form of morphing to normalize feature-spacing and emphasize the information provided by other characteristics such as the shape or color of individual features (Zhao et al., 2003). Burton, Jenkins, Hancock and White (2005) model face recognition by human observers using an algorithm that partially removes the information provided by feature-spacing. Given the number of ways that faces could be recognized, which do not involve judgments of feature-spacing, it seems logical to ask if our sensitivity to feature-spacing is correlated with face identification accuracy. The notion that feature-spacing judgments support face identification implies that individuals who are sensitive to differences in IPD and NMD ought to identify faces more accurately than individuals who are not sensitive to such differences. This hypothesis was tested in Experiment 5, where observers from Experiments 3 and 4 identified high-contrast, unfamiliar faces in a lineup task

(Bruce et al., 1999).

4.2 Results

4.2.1 Weber fractions for feature-spacing in realistic conditions

During each feature separation discrimination task, observers were shown a sequence of two briefly presented faces, one of which had a smaller IPD (Experiments 1, 2, and 3), or a smaller NMD (Experiment 4). IPD and NMD are depicted in Figure 4.1. In Experiments 1 and 2, the two faces shown during a single trial subtended the same visual angle. Therefore, observers were asked to indicate whether the first or the second face possessed the smaller IPD. In Experiments 3 and 4, the two faces shown on a trial differed in visual angle by a factor of 2. Therefore, observers were asked to indicate whether IPD or NMD was *relatively* smaller in the first or the second face. In Experiments 1 and 2, the IPD of one face, designated the standard, was held constant at 83 pixels. In Experiment 3, the IPD of the standard face was either 83 or 166 pixels. In Experiment 4, the NMD of the standard face was either 20 or 40 pixels. The lengths of IPD and NMD in the standard faces of Experiments 3 and 4, respectively, correspond to the largest IPD and NMD among the the set of faces prior to size normalization (see Methods).

In each experiment, observers completed a block of trials at each of six viewing distances: 50, 59, 70, 85, 100 and 120 cm. The order of viewing distance was randomized for each observer. For each block of trials, the difference in IPD or NMD was varied using the method of constant stimuli. Eight different levels of spacing differences that spanned the threshold range were used. The proportion of correct responses for each difference was recorded, and Weibull functions were fit to the data. Discrimination threshold was defined as the 75% correct point on the resulting psychometric function. Each threshold was converted into units of visual angle.

For Experiments 3 and 4, thresholds were measured at all viewing distances

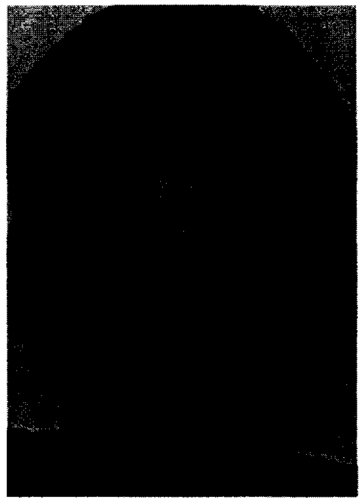


Figure 4.1: The facial distances used in these experiments. Inter-pupillary distance or IPD (1); Eye-to-mouth distance (2); Nose-to-mouth distance or NMD (3); Eye-to-nose distance (4).

on each of six days of testing. All observers exhibited some degree of learning effects. Therefore, thresholds were averaged across the last few days after learning subsided. The final result for each observer, in each experiment, was a single function relating IPD (or NMD) discrimination threshold to the length of IPD (or NMD) in the standard face.

If Weber's Law was valid for our conditions, then IPD (or NMD) discrimination thresholds should be proportional to IPD (or NMD) in the standard face. To evaluate this idea, we measured the goodness-of-fit, as indexed by R^2, of a line fit to the discrimination data. Examples of best-fitting lines fit to discrimination data from each experiment are shown in Figure 4.3. To obtain a single Weber fraction for each observer, we calculated the average ratio between IPD (or NMD) threshold and size of standard IPD (or NMD) across all viewing distances. For the rest of this paper, with the exception of Facial Proportions section, performance in Experiments 1-4 is analyzed with reference to these Weber fractions (shown in Figure 4.4), rather than the size-specific IPD (and NMD) thresholds (as in Figure 4.3).

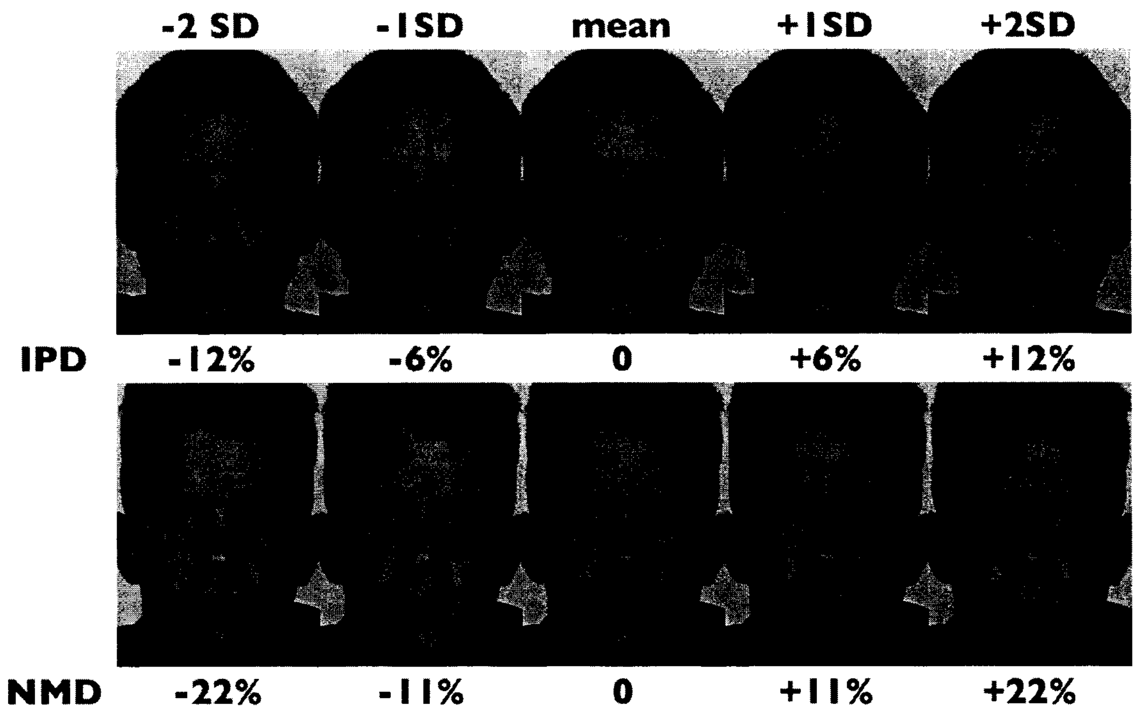


Figure 4.2: The range of IPD (first row) and NMD in real faces (second row), based on facial anthropometry data from Young (1993). The face in the center has the mean value for IPD or NMD. One standard deviation above and below the mean, IPD changes by about 6% of the mean distance, while NMD changes by about 11% of the mean.

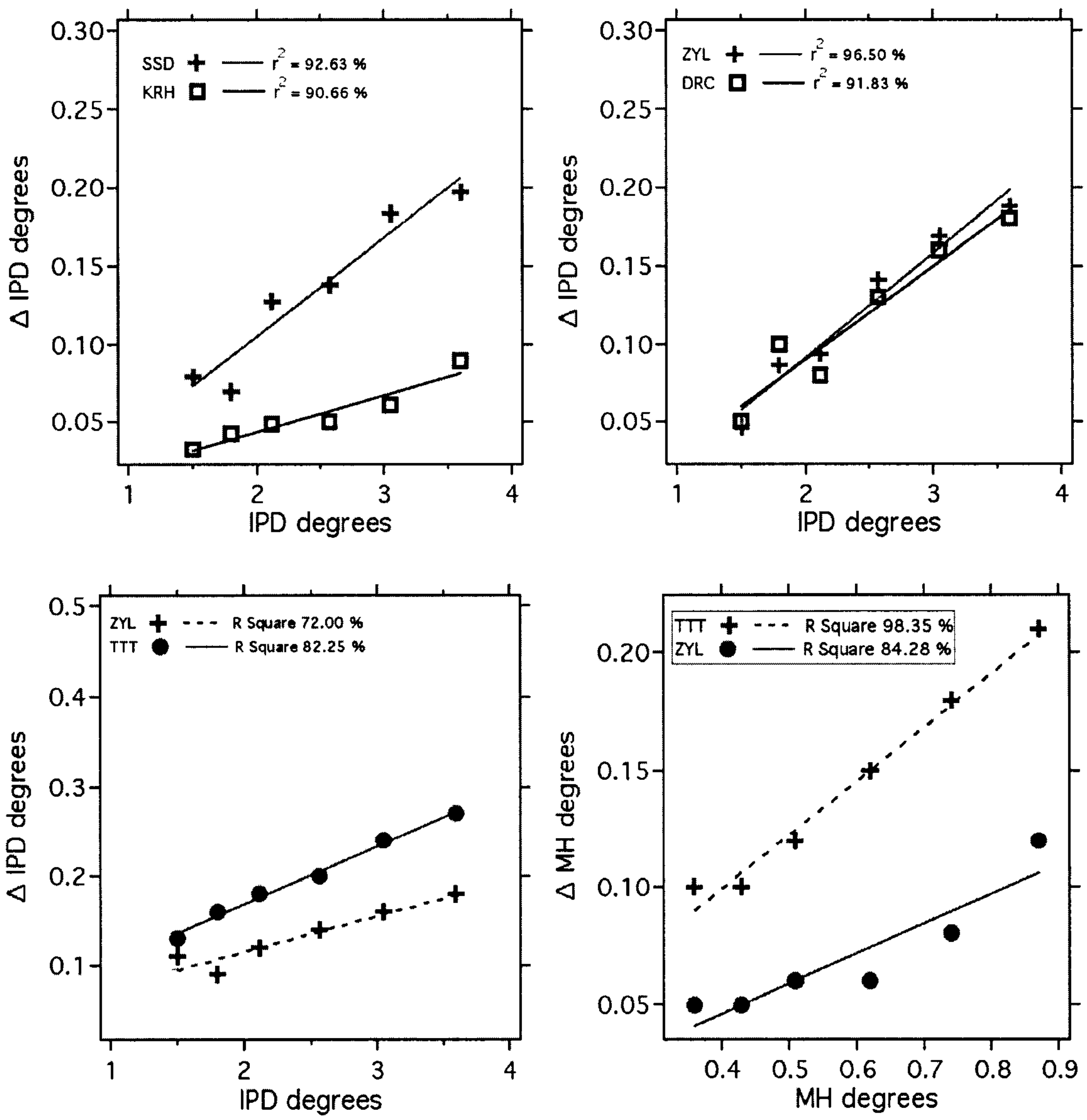


Figure 4.3: Example Weber functions for IPD discrimination in Experiments 1 (top left), 2 (top right), and 3 (bottom left); and for NMD in Experiment 4 (bottom right).

4.2.1.1 Experiment 1

All the faces were "large" (as defined in the Methods section). Although there were 8 different facial identities, randomly presented with equal frequency within each block of trials, pairs of faces were always from the same person. Therefore, the pair of faces shown on each trial differed only in IPD. Four observers were tested and Weber's law gave an excellent fit to each of their results (see Table 4.1). The mean R^2 of the linear fit was 89.5% with a minimum of 85%. The mean Weber fraction across our four observers was 4.8% ($\pm$1.61 SD; see Figure 4.4).

The threshold Weber fraction for each observer was converted to a P.E.D using the methods described in the section *Calculation of PED.* PEDs ranged from 40% to 80%; the mean value was 60% (see Figure 4.5). However, even the highest PED in this experiment is quite low. Recall that PED is the probability of using IPD alone to discriminate two randomly selected faces at *threshold performance.* In our experiment, threshold performance was 75% correct. Therefore, even an observer with a PED of 80% would still make a substantial number of errors on the occasions in which the IPD difference was above threshold. Our analyses suggest that IPD *by itself* cannot be used to reliably discriminate two faces chosen randomly from the population of Caucasian faces.

4.2.1.2 Experiment 2

Five observers were shown pairs of sequentially presented faces, and identified the face with the smaller IPD. All the faces shown were "large" (as defined in the Methods). Unlike Experiment 1, however, pairs of faces were always from different persons. Although the pair of faces shown on each trial differed in both IPD and identity, observers were instructed to judge only the change in IPD. Except for the potentially distracting variation in facial identity, Experiment 2 was the same as Experiment 1.

Weber's law gave an excellent fit to the results of all 5 of the observers in this experiment: the mean R^2 of the linear fit was 89% with a minimum of 73% (see Table 4.1). The mean Weber fraction was 4.64% ($\pm$1.02 SD; see Figure 4.4). In order to test the hypothesis that variations in facial identity increase Weber fractions for IPD discrimination, a two-sample t-test, assuming unequal variances, was used to evaluate the difference between Weber fractions measured in Experiments 1 and 2. Thresholds in the two experiments did not differ significantly (t=0.13, df=4.4, p=0.90), suggesting that variations in facial identity did not affect the discrimination of differences in IPD.

PED values were approximately the same as those obtained in Experiment 1 (see Figure 4.5).

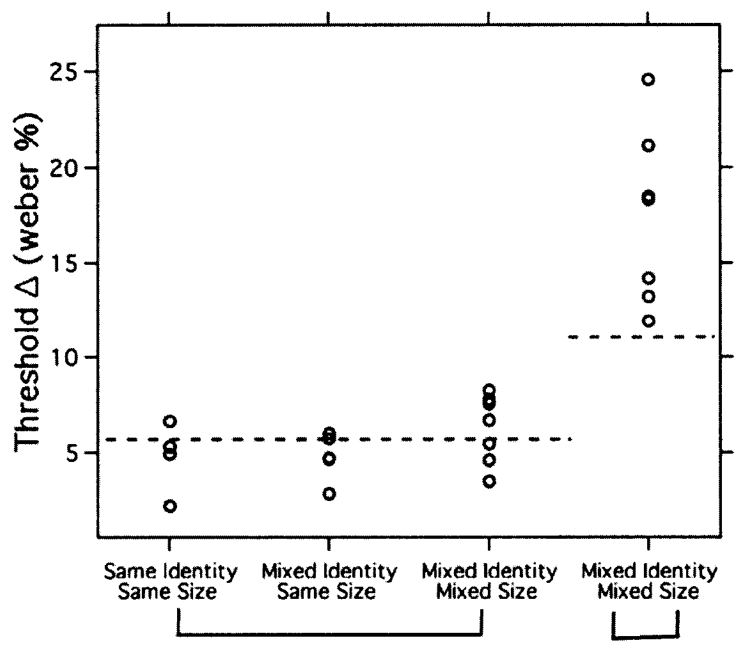


Figure 4.4: Distribution of Weber fractions across observers in (from left to right) Experiments 1 to 4. Dotted lines correspond to the coefficient of variation of either IPD (black line) or NMD (red line) in the real population of Caucasian faces.

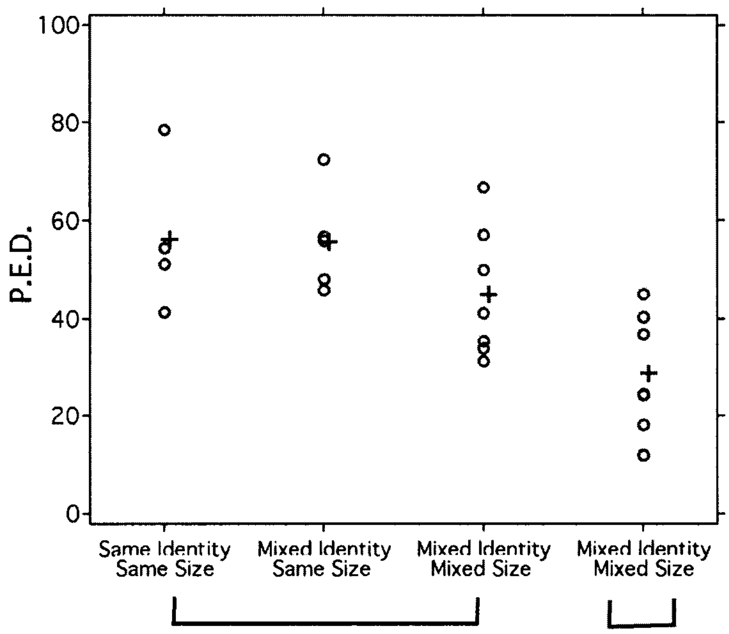


Figure 4.5: Distribution of PED across observers in Experiments 1 to 4, going from the far left column to the far right column. Black crosses indicate the mean PED for the corresponding distribution.

4.2.1.3 Experiment 3

Seven observers were shown pairs of sequentially presented faces, and identified the face with the smaller IPD. As in Experiment 2, the faces presented on each trial differed in identity. However, in addition, the faces on each trial also differed in size. Although the pair of faces shown on each trial now differed in 3 ways (IPD, identity, and size), the observers were instructed to judge only the change in IPD. Except for the potentially distracting variation in face size, Experiment 3 was the same as Experiment 2.

Thresholds were estimated by first rescaling each target face so that it was the same size as the standard, and then calculating the difference between the IPDs of the rescaled target and standard faces. (We discuss the validity of the rescaling in the section *Facial Proportions*). Weber's law gave a reasonable fit to the results of 5 of the 7 observers in this experiment (see Table 4.1); for these 5 observers, the mean R^2 of the linear fit was 78% with a minimum of 71%. The two observers giving bad fits had an R^2 of 63% and 22%. In the section entitled *Facial Proportions*, we discuss why these observers deviated so significantly from Weber's law. The mean Weber fraction across our seven observers was 6.28% ($\pm$1.65 SD; see Figure 4.4). In order to test the hypothesis that combined variations in facial identity and size increase Weber fractions for IPD discrimination, a two-sample t-test, assuming unequal variances, was used to evaluate the differences between Weber fractions measured in Experiments 1 and 3. This test was not significant (t = 1.29, df = 6.13, p = 0.24), suggesting that combined variations in facial identity and size do not affect the discrimination of differences in IPD.

The average value of PED was less than 50%, which was slightly lower than values obtained in the previous two experiments (see Figure 4.5).

We also tested the hypothesis that additional variation in the size of faces, on top of an existing variation in facial identity, increased Weber fractions for IPD discrimination. Experiments 2 and 3 can be compared using a paired t-test because four of the same observers performed in both. The scatterplot

	Experiment			
Observer	1	2	3	4
BRB	85	87	63 (78)	96
GRG	—	—	82 (71)	91
KRH	91	73	22 (82)	89
NIC	—	96	72	80
SSD	93	—	84	97
TTT	89	—	82	98
ZYL	—	97	72 (27)	84
DRC	—	92	—	—
Mean	89.5	89.0	68.1 (70.8)	90.1

Table 4.1: Percent R^2 for fits of Weber's Law model to thresholds from Experiments 1-4. Bracketed values indicate R^2 after original stimulus were recoded (see text for details).

in Figure 4.6 shows that there is a consistent increase in Weber fraction in Experiment 3, compared to Experiment 2. However, a paired t-test reveals that this difference was not significant ($t = -2.4$, $df = 3$, $p = 0.10$). This result suggests that IPD discrimination is not significantly worse when observers must compare faces that differ in visual angle.

4.2.1.4 Experiment 4

In Experiment 4, observers discriminated the distance between the bottom of the nose and the midpoint of the fissure between closed lips (i.e., distance 3 in Figure 4.1). In this experiment, the faces always differed in identity and size, allowing us to compare the sensitivity of NMD discrimination with that of IPD thresholds measured in Experiment 3.

Thresholds were estimated by first rescaling each target face so that it was the same size as the standard, and then calculating the difference between the NMDs of the rescaled target and standard faces. (We discuss the validity of the rescaling in the section *Facial Proportions*). Weber's law gave an excellent fit to the results of all 7 of the observers in this experiment (see Table 4.1).

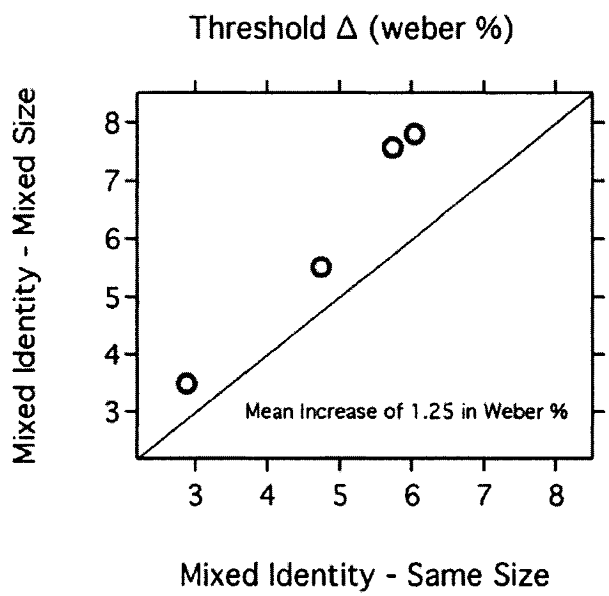


Figure 4.6: Scatterplot of Weber fractions IPD in Experiments 2 (x axis) against those for Experiment 3 (y axis). Points above the line indicate a cost of judging changes in IPD in object-relative units compared to image units. On average, Weber fractions increased 25%.

The mean R^2 of the linear fit was 90% with a minimum of 80%. Weber's law, in fact, gave a slightly better fit to NMD judgments than to IPD judgments measured in Experiment 1, when observers compared IPD in faces that did not differ in identity or size. The mean Weber fraction across our seven observers was 17.41% ($\pm$4.23 SD; see Figure 4.4). Weber fractions for NMD discriminations were much higher than those for the IPD discriminations measured in Experiment 3 ($t = 9.387$, $df = 6$, $p < .0001$).

PED values for NMD discrimination were very low, less than 40% on average (see Figure 4.5). Therefore, NMD is a less useful cue than IPD for face identification in a natural context. However, neither NMD nor IPD by themselves would be particularly useful for face identification.

A paired t-test of PED calculated from IPD and NMD thresholds measured in Experiments 3 and 4 was significant, indicating that PED was significantly lower for NMD ($t = 6.24$, $df = 6$, $p = 0.0008$). Despite the fact that NMD has a higher coefficient of variation than IPD in the population of Caucasian faces, therefore, our analyses indicate that observers would be less able to use differences in NMD than IPD to discriminate Caucasian faces. This result is not consistent with the idea that sensitivity to differences in feature-spacing is positively related to the coefficient of variation of feature-spacing in real faces.

4.2.2 Facial Proportions

Observers in Experiment 3 were required to judge IPD on an object-relative scale, rather than the absolute length of IPD. How might observers obtain a scale-invariant metric of IPD? Thresholds were estimated by first rescaling each target face so that it was the same size as the standard face, and then calculating the difference between the IPDs of the rescaled target face and the standard face. This analysis is equivalent to one in which IPD is normalized by an inter-feature distance or feature that is invariant across identity. Therefore, we assume that our observers were able to use a person-invariant distance as a reference to convert IPD into a scale-invariant proportion.

For the specific task and stimuli used in Experiment 3, iris diameter may have served as a valid reference because it was the same for all of the standard faces. Also, it is interesting to note that the diameter of the human iris has a low coefficient of variation (i.e., 3.5-4%; Martin & Holden, 1982; Theodorff & Lowther, 1990). Against the bright-white sclera, the iris may also constitute the region of highest local contrast within the human face. This leaves us with the question of whether human observers are able to use the iris diameter, or some other facial distance, to normalize IPD. We can only infer the usage of a particular reference by returning to the raw data and recoding our IPD difference levels as changes in the ratio of IPD to the proposed reference distance. Recoding the IPD-difference as a change in IPD-ratio may actually change the 'correct answer' on a given trial if we use a reference distance that is, unlike the iris, variable across facial identity. In turn, stimulus recoding may change may shape of the psychometric function relating stimulus level to percent correct and, by extension, the relative pattern of threshold stimulus level across viewing distance.

Recall that, in Experiment 3, Weber's Law provided a good fit to the IPD discrimination thresholds for all but 2 of the 7 observers. Those 2 observers, KRH and BRB, may have differed from the other 5 because they used a different reference against which to judge IPD, presumably something less stable across facial identity. If that is the case then it would not be appropriate to apply Weber's Law to their thresholds without first recoding their data according to whichever facial reference they were actually using. Barton, Zhao and Keenan (2003) provide evidence for an interaction between judgments of IPD and judgments of the distance between the eyes and mouth. Therefore, we determined if the thresholds for some of our observers in Experiment 3 could be better fit by assuming they used eye-to-mouth height as a reference for judging IPD. We recoded IPD-difference as IPR (the log ratio of IPD divided by eye-to-mouth height; see distance 2 in Figure 4.1). To assess the improvement to psychometric function fits made by recoding stimulus levels, we used a two-sided signed rank test to compare the R^2 values for those fits before and after recoding (using a criterion p=0.05). Recoding the data improved the fits of the psychometric functions significantly in observers BRB, KRH, GRG and ZYL (see Figure 4.7), which suggests that these observers may have used eye-to-mouth height — or another metric that was correlated with it — to normalize IPD when comparing faces that differed in visual angle. Interestingly, observers BRB and KRH were the two observers who failed to demonstrate Weber's Law in Experiment 1. However, if we now assume that these two observers had judged IPR instead of IPD-difference, BRB's R^2 for Weber's law improves from 63% to 78%, and KRH's improves nearly 4-fold from 22% to 82% (see Table 4.1). GRG's R^2 dropped from 82% to 71%, which is still a reasonable fit to Weber's Law. However, ZYL's R^2 dropped from 72% to 27%. Once this inferred difference in strategy is taken into account, Weber's law provides a good account of behavior of 6 instead of 5 of the 7 observers we tested. Therefore, whether an inferred difference in strategy is taken into account or not, the majority of observers obeyed Weber's Law for IPD.

Observers in Experiment 4 were required to judge NMD on an object-relative scale. To examine whether observers coded NMD relative to some other distance, we recoded the independent variable from NMD to the ratio of NMD divided by eye-to-nose height, in the same way that we recoded IPD values previously. We then determined the improvement to psychometric function fits by comparing R^2 for those fits before and after recoding using a two-sided rank sign test (using a criterion p=0.05). The psychometric functions for all observers except for BRB demonstrated a significant gain in R^2 by assuming observers judged NMD relative to eye-to-nose height (i.e., distance 4 in Figure 4.1). This result may be a function of how we normalized the standard faces for this experiment. Unlike Experiment 3, the faces used in Experiment 4 were adjusted to have a common standard NMD, and (consequently) had different iris diameters. Therefore, it would not have been optimal for our observers to have normalized NMD by iris diameter in our task, or any other typically stable facial feature. Instead, the reference that observers in Experiment 4 appear

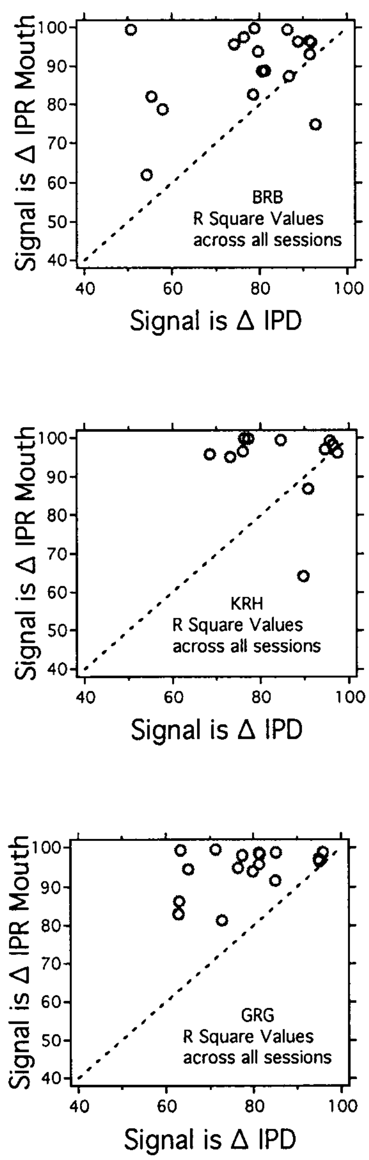


Figure 4.7: Scatter plots of fits of psychometric functions to IPD discrimination data collected in each condition from three observers. The x and y axes show $\%R^2$ for the fits to the original data and the recoded data, respectively. The recoded independent variable, IPR, was defined as the ratio of IPD divided by the eye-to-mouth distance (i.e., distance 2 in Figure 4.1). The dashed line has a slope of 1 and intercept of zero. The points generally fall above the dashed line, indicating that the psychometric functions fit the data from these observers better when the independent variable was defined as IPR rather than IPD.

to have been using was perhaps even more unstable. Eye-to-nose heights differed by as much as 19% across the standard faces used in Experiment 4. And yet, to their detriment, all of our observers appeared to be using eye-to-nose height as a reference distance anyway. The use of an unstable refrence may be one reason why thresholds were significantly higher for NMD compared to IPD. Nevertheless, our chief concern in this section is to determine if inferred variations in feature-spacing strategy can be accounted for to improve the fit of Weber's Law to observer thresholds. In the case of NMD judgments, the fit of Weber's Law is already very good for all of our observers (see Table 4.1). This raises the possibility that the fit of Weber's Law will worsen once inferred variations in strategy are taken into account. However, this is not the case. Using the stimulus-recoded thresholds, we recalculated the fit of Weber's Law for all 6 of the observers whose thresholds were better estimated by assuming a change in strategy. R^2 values for the fit of Weber's Law were reduced in three of the observers (from 89.65% to 79.94% in KRH, from 97.58% to 76.74% in SHD, and from 98.23% to 87.51% in TTT), stayed the same in one observer (from 79.86% to 79.26%), and increased in two observers (84.48% to 91.5% in ZYL, and 90.74% to 94.52% in GRG). We also compared the R^2 values for Weber's Law fits before and after re-coding using a paired t-test; the difference in R^2 was not significant ($t = 1.54$, $df = 6$, $p = 0.175$). Therefore, whether an inferred difference in strategy is taken into account or not, all of the observers obeyed Weber's Law for NMD discrimination, including the observer who did not obey Weber's Law for IPD in Experiment 3 (ZYL).

4.2.3 Association between spatial discrimination and face identification

Does face identification rely on the same sorts of processes that are used to support feature-spacing judgments? To determine if feature-spacing and face identification are related, all of the observers who participated in Experiments 3 and 4 also performed a face identification task (designed by Bruce et al., 1999; see *Methods* for details). On each trial observers were shown a target

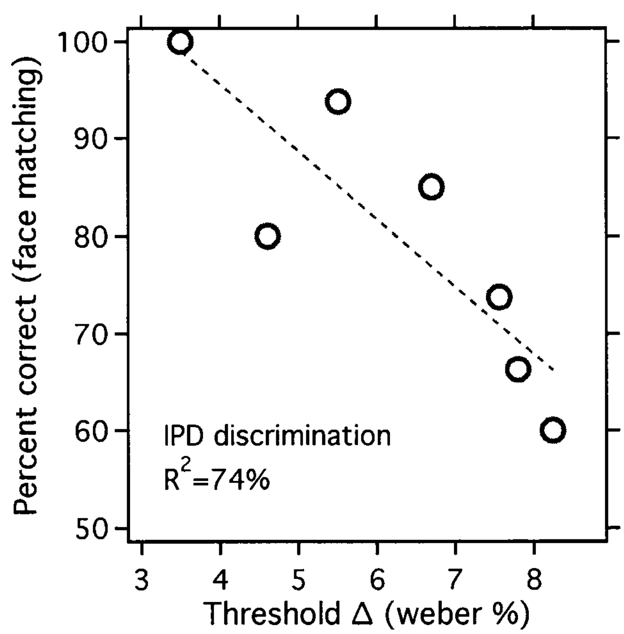


Figure 4.8: The relation between response accuracy in the lineup task and IPD discrimination threshold. Each point represents the results from a single observer. The dashed line is the regression line. The correlation between the two measures was -0.86 ($R^2 = 0.74$).

face together with a lineup of ten other faces. All of the faces were unfamiliar. The observer's task was to determine if the target face was present in the lineup and, if so, which face matched the target. Faces were high-contrast and unedited (i.e., facial contour, hair, and internal features were unchanged). The target face was present in the lineup on 50% of the trials. When the target face was present, it was depicted by another photograph taken on the same day as the target photo. Therefore, the observer could not perform exact "picture-matching" to produce a correct response. Percentage of correct responses was measured for all of the observers who performed in Experiments 3 and 4.

Just as Bruce et al. (1999) demonstrated in their original study, individual observers varied substantially in face-matching accuracy, with response accuracy ranging from 57 to 94 percent correct ranging from 60 to 100 percent correct. We wanted to know if sensitivity to feature-spacing, measured with faces that differed in size and identity, contributed to this variation in face identification accuracy. Therefore, we correlated Weber fractions for IPD in Experiment 3, and those for NMD in Experiment 4 with the accuracy of face identification[1]. Scatterplots relating the accuracy of face identification to threshold Weber fractions for IPD and NMD are shown in Figures 4.8 and 4.9, respectively. As one can see, both IPD and NMD thresholds were highly predictive of face identification accuracy. In fact, threshold IPD accounted for 74% of the variance in identification accuracy ($F = 14.05$, $df = (1, 6)$, $p = .002$), and threshold NMD accounted for 81% of the variance in accuracy ($F = 20.9$, $df = (1, 6)$, $p < .001$).

Why are IPD and NMD discrimination thresholds correlated with accuracy in the lineup task? One possibility is that face identification in the lineup task is based, in part, on discriminating differences in certain aspects of facial geometry (e.g., IPD and/or NMD). However, another possibility is that per-

[1]We did not use the stimulus-recoded Weber fractions described in Facial Proportions because that would have inflated the performance of observers who may have used a relatively poor strategy for judging feature-spacing, compared to those observers who did not require a recoding for improved threshold fits.

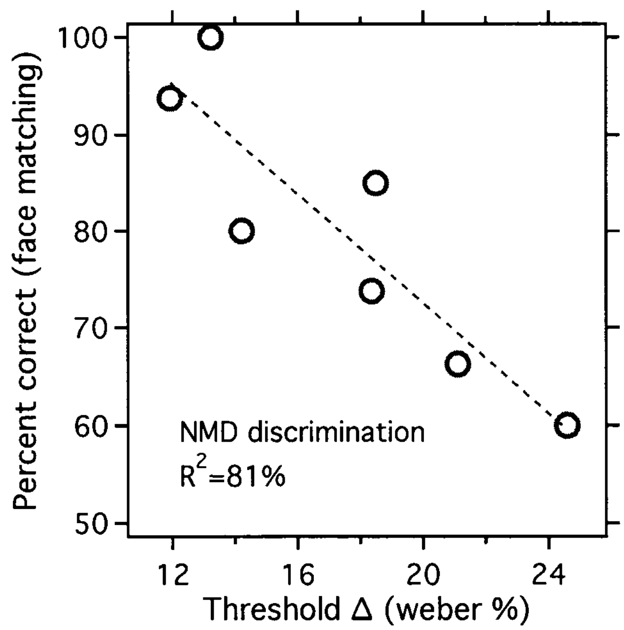


Figure 4.9: The relation between response accuracy in the lineup task and NMD discrimination threshold. Each point represents the results from a single observer. The dashed line is the regression line. The correlation between the two measures was -0.90 ($R^2 = 0.81$).

formance on both sets of tasks is influenced similarly by high-level factors such as motivation, vigilance, the ability to focus attention on one variable and/or ignore variation on irrelevant dimensions, etc. To examine the influence of such factors, we compared performance in the lineup task to thresholds in a face-contrast discrimination task. Contrast discrimination thresholds were measured with the same faces and methods used to measure IPD discrimination thresholds in Experiments 1 and 3, except that IPD was held constant and faces varied only in terms of overall contrast. We assumed that the effects of high-level factors such as motivation, vigilance, and the ability to focus attention on task-relevant variables, would be similar in the IPD, NMD, and contrast discrimination tasks. However, unlike the IPD and NMD discrimination tasks, the contrast discrimination task does not require observers to accurately encode the spatial distribution of features within a face. Furthermore, it is unlikely that face identification is based on noting differences in overall contrast. For these reasons, we would argue that a significant correlation between face-contrast discrimination and face identification would primarily reflect the common influence of higher-level factors. Due to attrition, only 6 of the 7 observers who participated in Experiments 3-5 participated in the contrast discrimination experiment. As shown in Figure 4.10, contrast discrimination thresholds were significantly negatively correlated with accuracy in the lineup task ($R^2 = 91\%$, $F = 42.3$, $df = (1, 5)$, $p < 0.001$).

The results of the contrast discrimination experiment suggest that the correlations between performance in the lineup task and IPD and NMD discrimination thresholds reflect, at least in part, the common influence of high-level factors, rather than a reliance on the same low-level mechanisms. It remains possible, however, that the correlations also reflect the common influence of mechanisms that encode the positions of features within a face. To test this idea, we examined whether IPD and NMD thresholds were correlated with accuracy in the lineup task after controlling for the statistical association between accuracy and contrast discrimination thresholds. Thresholds in the IPD and NMD discrimination tasks were averaged to create a composite score that

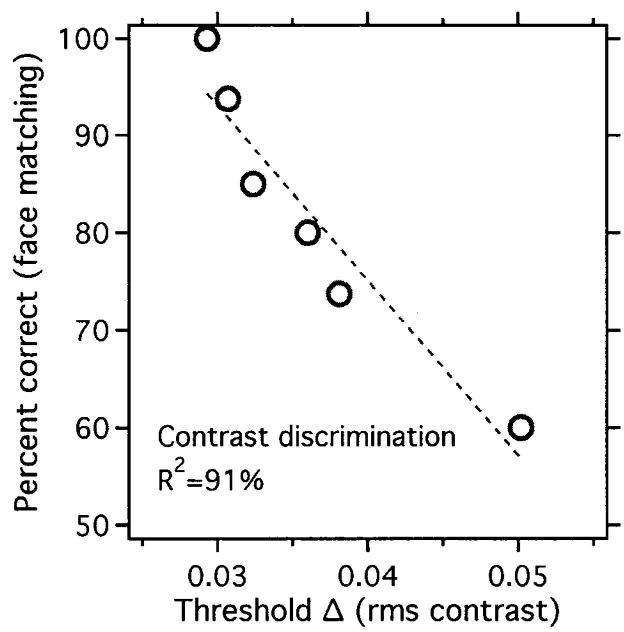


Figure 4.10: The relation between response accuracy in the lineup task and contrast discrimination threshold. Each point represents the results from a single observer. The dashed line is the regression line. The correlation between the two measures was -0.96 ($R^2 = 0.91$).

reflected each subject's ability to encode the spatial arrangement of facial features. Next, we z-transformed the composite score, as well as contrast discrimination thresholds and response accuracy in the lineup task. Finally, we used hierarchical multiple regression to calculate the effect of adding the composite score to a model that already included contrast discrimination threshold as a predictor of lineup response accuracy. This two factor model provided an extremely good fit to the lineup accuracy data (R^2 = 93%, F = 19.97, df = (2, 3), p = .018). The sequential sums of squares for each term in the model are presented in Table 4.2. Note that the average of the IPD/NMD thresholds, labeled as IPD/NMD in the table, was not significantly associated with accuracy in the lineup task after controlling for the statistical association between accuracy and contrast discrimination thresholds. This result indicates that performance in the IPD and NMD tasks, on the one hand, and performance on the contrast discrimination task, on the other, account for overlapping portions of the variance of the response accuracy scores. Furthermore, this result suggests that the high correlation between face identification accuracy and IPD and NMD discrimination thresholds may reflect the influence of common, high-level factors on task performance.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
z(contrast)	1	4.57	4.57	39.24	0.0082
z(IPD/NMD)	1	0.08	0.08	0.70	0.4656
Residuals	3	0.35	0.12		

Table 4.2: Sequential sums of squares for two terms in a linear model that predicted accuracy in the face lineup task from contrast discrimination thresholds (contrast) and the average of IPD and NMD thresholds (IPD/NMD). All variables were z-transformed.

4.3 Discussion

Many researchers have claimed that the discrimination of feature-spacing, in general, is an important skill used during face recognition (Diamond & Carey, 1986; Gauthier, Williams, Tarr & Tanaka, 1998; Kemp, McManus, &

Pigott, 1990; Le Grand, Mondloch, Maurer, and Brent, 2001; 2003; Leder & Bruce, 2000; Leder, Candrian, Huber, and Bruce, 2001; Maurer, Le Grand, & Mondloch, 2002; Rhodes, 1988; Searcy & Bartlett, 1996). The current study took two complementary approaches to examine whether the discrimination of IPD and NMD plays an important role in face recognition. First, we measured the extent to which variations in IPD and NMD can support face identification in a natural context. In order to accomplish this, we made novel use of the statistics of real faces, comparing the available ranges of IPD and NMD against observer thresholds for the discrimination of these cues. Second, we determined if face identification is supported, in part, by an ability to discriminate the geometry of faces, like differences in IPD and NMD. In order to answer this question we took advantage of the large individual differences in accuracy during a realistic face identification task (Bruce et al., 1999), much of which should be explained by a thorough understanding of the different skills required by this demanding task.

Discrimination thresholds for IPD and NMD were approximately invariant to changes in facial identity and size. This result suggests that observers possess the kind of pattern-invariant processing that would be necessary if feature-spacing is to make any contribution to face identification in realistic conditions. However, our comparison of IPD and NMD discrimination thresholds against the limited amount of natural variation in these cues demonstrated that neither IPD nor NMD would be useful for face recognition in a natural context. Of course, the relative utility of different types of facial cues is a separate issue from the relative importance of different types of perceptual skills. It is possible that the skills underlying the discrimination of IPD and NMD are especially important for face identification. However, we found that performance in a face identification task is not correlated with feature-separation thresholds, once the common influence of high-level factors is controlled for. This result suggests that the skills underlying judgments of facial feature-spacing did not contribute significantly to accuracy in the face identification task.

One of the most important goals of face identification research is to understand what aspects of the human face are most important for visual identification. If face identification is not supported by the judgment of feature-spacing, how should future research proceed? As we mentioned earlier, the importance of a facial cue is a separate matter from how we process that cue. Even if the skills underlying judgments of feature-spacing are not particularly important for face identification, feature-spacing cues that are significantly more variable in the population than IPD and NMD might still be important for identification. However, choosing a candidate facial-distance appears to be a daunting task because there are a great many facial-distances one could evaluate. Fortunately, facial statistics can help in this regard, by allowing us to generate reasonable hypotheses about which facial-distances should be used by human observers. For example, one can rank order the facial anthropometry data of Farkas (1981) by the coefficient of variation for each feature-distance. Near the top of this list is the distance between the eyebrow and the iris (EID), which has a coefficient of variation of about $\approx 17\%$, which is much larger than the coefficients of variation for IPD and NMD. If feature-spacing is important for face recognition, then one would expect discrimination thresholds to be low for EID. One would also expect the PED to be higher for EID than it is for both IPD and NMD. This hypothesis is consistent with the classification image results of Sekuler et al. (2004), who demonstrated that observers identify faces by using information in only a small region of the face that includes the eyes and eyebrows. It is possible that these classification image results reflect a strategy that calculates variation in EID. It is also possible that the region around the eyes and eyebrows is important for multiple reasons, one of which is the strong variation in EID across the population of faces. For example, there may be something particularly salient about the shapes and textures around the area that includes the eyes and brows, like their relatively higher local contrast. In general, future studies will have much to gain by considering individual differences in face identification ability, the ecological constraints on face recognition, and the statistics of natural facial variations.

4.4 Methods

4.4.1 Observers

The subjects were eight (3 female and 5 male) experienced psychophysical observers with normal or corrected-to-normal visual acuity.

4.4.2 Stimuli

4.4.2.1 Photography

The stimuli in the IPD and NMD discrimination experiments were constructed from digital photographs of eight Caucasian faces (4 male and 4 female) of undergraduate students between 18 and 21 years old (average age = 20 years). The photographs were taken from a frontal view in standard lighting conditions, with the camera approximately the same distance from each subject. Camera resolution was 1800 x 2400 pixels (RGB, 8-bits per channel). The original photographs were taken by Lisa DeBruine (http://www.faceresearch.org/).

4.4.2.2 Position & size normalization

The photographs were processed with Adobe Photoshop 7.0 before being used in the experiment. First, images were rotated (on average 3 degrees) until the chin and apex of the forehead were aligned with the vertical axis. Second, image size was normalized by either inter-pupillary distance or nose-to-mouth distance, depending on whether inter-pupillary distance (IPD) or nose-to-mouth distance (NMD) discrimination thresholds were measured, respectively. For IPD experiments, inter-pupillary distance was exactly 83 pixels for all faces; for the NMD experiment, nose-to-mouth distance was exactly 20 pixels for all faces. Then, the images were cropped to be 314 x 226 pixels. Image cropping followed rules to ensure a standard alignment of the face within the resized image: the lowest point of the chin aligned with the bottom of the image, and the sides of the image were equal distances away from the sides of the face.

4.4.2.3 Inter-pupillary distance

For each of the eight facial images, another eight were created which differed from the original only in the distance between the eyes (number 1 in Figure 4.1). Elliptical regions were selected around the eyes, just large enough to encompass the lower and upper folds of the eyes, along with the inner and outer corners; for faces with long eyelashes, the elliptical regions had to be slightly larger. The elliptical regions were moved inward by equal increments, and the gaps created by these displacements were filled in by superimposing the edited image over a copy of the original. Discontinuities in skin-tone and other artifacts of displacement were removed by blurring. Eight new versions of each original face had an IPD that was smaller than the original by a total of 2, 4, 6, 8, 10, 12, 14 or 16 pixels. Finally, the images were converted to 8-bit gray-scale and saved as uncompressed bitmaps.

4.4.2.4 Nose-to-mouth height (NMD)

For each of the eight facial images, another eight were created which differed from the original only in the distance between the bottom of the nose and the middle of the labial fissure (number 3 in Figure 4.1). These images were created in the same manner as the IPD stimuli described above. Eight new versions of each original face had an NMD that was smaller than the original by a total of 1, 2, 3, 4, 5, 6, 7 or 8 pixels. Finally, the images were converted to 8-bit gray-scale and saved as uncompressed bitmaps.

4.4.2.5 Image size & contrast

Two different image sizes were used in this study: Large images were 314 x 226 pixels, and small images were 157 x 113 pixels. The set of smaller images were created by resizing the larger set of images in MATLAB, using the method of bicubic interpolation available in the Image Processing Toolbox. All images were normalized to have an rms contrast of 0.35 when displayed on the monitor.

4.4.3 Discrimination of feature separation

All stimuli were viewed on a 21-inch Apple Studio Display monitor resolution with a resolution of 1024 x 768, and a frame-rate of 85 Hz. At this resolution, the displayable portion of the monitor measured 38.7 cm wide. Average screen luminance was set to 30 cd/m^2. Observers viewed the monitor from six different distances: 50, 59, 70, 85, 100, and 120 centimeters.

On each trial, two faces were presented either 256 pixels to the left or right of the center of the display. Changing the location of the face within a trial prevented observers from perceiving apparent motion of as a result of changing the positions of facial features. To reduce spatial uncertainty, rectangular frames (4 pixels wide, -0.5 contrast) were drawn around the two locations and remained visible for the duration of the experiment. The location of the first face was randomized across trials, with an equal number of first faces appearing to the left of fixation as to the right. The location of the first face was cued by a circular disc (radius=10 pixels; contrast = -0.5) that was presented for 200 ms. Observers had 1s to fixate the cued location. The first face was then displayed for 200 ms, followed by a 500 ms blank screen, during which time the observers were instructed to fixate the other side of the display. Finally, the second stimulus was presented for 200 ms. After the second face was extinguished, the display remained blank until the observer made a decision. Auditory feedback, in the form of 600 and 200 Hz tones, was presented after correct and incorrect responses, respectively. Observers were told to respond as accurately as possible, and to take as long as necessary to make their decision.

Observers in our task were asked to decide between two temporal intervals, but we were well aware that observers might still confuse temporal order with spatial location (e.g., the Simon effect). To ensure that our observers were not making this mistake, observers performed 20-30 practice trials at a viewing distance of 70 cm prior to beginning their first block of test trials. During practice trials, a research assistant sat nearby to confirm that the observer was comfortable with the task. All observers were confident that they were doing the task correctly.

The difference in feature separation (i.e., $\triangle IPD$ or $\triangle NMD$) on each trial was randomly selected from eight values that spanned the threshold range. During each block, each value of $\triangle IPD$ (or $\triangle NMD$) was presented on 24 trials.

4.4.4 Calculation of PED

The distribution of IPD among adult Caucasian faces is well described by a Gaussian density function (Dodgson, 2004). We assumed that NMD is also well described by a Gaussian function. Therefore, the distance between facial features, D, is distributed as a normal random variable.

$$D \sim N(\mu, \sigma)$$

where $\sigma = c\mu$ and c is the coefficient of variation. If we randomly sample two faces from the population, we get two distances, d_j and d_i. A new random variable can be calculated by subtracting the second feature distance from the first:

$$\triangle D = d_i - d_j$$

which is distributed as

$$\triangle D \sim N(0, \sigma\sqrt{2})$$

Judgments of IPD and NMD follow Weber's Law: the just-discriminable difference is a constant proportion of the larger facial distance. Therefore, we normalize our random variable $\triangle D$ by the distance of the larger facial distance:

$$F = \frac{\triangle D}{\max(d_i, d_j)}$$

F can now be directly compared to thresholds (expressed as Weber fractions). We do not know how F is distributed, so we used a random number generator to sample 106 pairs of d_j and d_i, and then used these values to approximate the probability distribution of F. PED then is simply the probability that F is greater than, or equal to, an observer's threshold:

$$\text{PED} = P(F \geq w)$$

where w is the threshold Weber fraction for a particular observer in a particular feature-distance experiment.

4.4.5 Contrast discrimination

The stimuli and procedure used in the face-contrast discrimination experiment were almost identical to those used in Experiment 3. Pairs of faces always had the standard level of IPD. Instead of differing by feature-spacing, the standard and comparison faces differed in their overall rms contrast. The observer's task was to decide which of the two faces had the lower contrast. The standard face always had an rms contrast of 0.35 while the comparison face always had a contrast value that was lower than 0.35. The contrast difference was varied according to two interleaved staircases, one following the 2-down/1-up rule and the other following a 3-down/1-up rule. Each staircase continued for a total of 120 trials. The proportion of correct responses for each level of contrast difference was recorded, and Weibull functions were fit to the data using a maximum likelihood procedure. Discrimination threshold was defined as the 75% correct point on the resulting psychometric function.

4.4.6 Face-matching

4.4.6.1 Viewing conditions

The stimuli were presented on a Sony Trinitron monitor in a dark room. Observers viewed stimuli from a distance of 45 cm; a chin rest was used to stabilize the viewing position. Monitor resolution was set to 2048 x 1536 (75 Hz), which subtended 46.6 x 35.7 degrees. Our experiments were programmed in MATLAB on a Power Mac G4, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). The monitor was calibrated to produce a linear range of luminance. The background luminance of the screen was set to 43 cd/m^2.

4.4.6.2 Stimuli

Stimuli consisted of 160 line-up displays (27.2 deg wide x 34.4 deg tall) created by Bruce et al. (1999). Each stimulus display was composed of a target face (4.5 deg x 7.6 deg) positioned at the top, and an array of 10 faces (each 3.8 deg x 6.4 deg) positioned at the bottom. Faces in the array were arranged in two rows of five faces: adjacent faces were separated by 0.5 deg. The bottom of the first row and the top of the second row were separated by 2.5 deg. A dark black (10 cd/m^2) numeral (0.6 deg x 0.6 deg) was displayed below each of the array faces, beginning with "1" for the first face in the top row and increasing left-to-right, top-to-bottom. The target face was centered horizontally in the display and spaced approximately 9.5 deg above the top row in the 10-item array. The area around each face was a uniform light-gray with a luminance of 80 cd/m^2. Each face was shown in its original high contrast and had an average local luminance of 40 cd/m^2. A selection panel (10.8 deg x 35.7 deg) was displayed to the left of the line-up display. On the selection panel was shown two columns of dark black (10 cd/m^2) numerals (1.5 deg x 1.5 deg), one column displaying the numerals 1 to 5 from top to bottom, and another column displaying 6 to 10. The word "absent" (9.2 deg x 1.5 deg) was displayed in dark black (10 cd/m^2) below the two columns of numerals.

For half of the line-up displays, the target face was included in the array of 10 faces. The target faces were generated by taking a single frame from a high-quality video camera, while the line-up faces were generated by using high-quality photographs taken hours after the target-images were obtained. Images of the target face at the top of a line-up image were therefore different from the image of the same face shown in the array. A similar frontal pose, lighting condition, and neutral expression was used for all photos but there were subtle differences in the target and array versions of the same face that prevented a pure picture matching strategy.

4.4.6.3 Participants

All but one of the observers who participated in the feature-spacing discrimination experiments also participated in the face-matching experiment.

4.4.6.4 Design

Each observer was randomly assigned to one of two stimulus sets: both sets contained 40 target-present line-ups and 40 target-absent lineups, but the target-present and target-absent versions of the same line-up were separated between the two sets.

4.4.6.5 Procedure

A research assistant read a standard set of instructions to the observer. During this time the observer was told that the experiment consisted of 80 trials, and that the target would be present on only half of the trials. He/she was also told that there was no time limit, and that they were to respond as accurately as possible. The observer adapted to the display for 2 minutes. During the first 1-3 trials, the research assistant sat with the observer to make sure that he/she understood the task.

At the beginning of each trial, the stimulus array was presented and a mouse cursor appeared in the center of the display. The observer's task was to examine the target face and try to find a match within the array of 10 faces. The observer indicated his/her decision by using the mouse to click on the face that matched the target. If the observer decided that the target face was not present in the array, then he/she clicked on the word "absent" that appeared below the array. If the cursor moved outside of the selection panel, it was automatically erased and redrawn at the center of the selection panel. This procedure prevented observers from using the cursor as a reference to measure the size of features or distances between features. After the observer made a decision, a blank screen of average luminance appeared for 200 ms, and then was replaced by the stimulus display for the next trial. The observer was not given feedback about the accuracy of the response.

All observers completed the experiment within approximately 25 minutes. Percent correct was the dependent variable. The correct response on a target-absent trial is to respond "absent", while the correct response on a target-present trial is to choose the matching face from the array of 10 faces. Therefore, percentage correct at chance performance was 1/11.

4.5 References

Barton, J.J., Keenan, J.P., & Bass, T. (2001). Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. British Journal of Psychology, 92 Part 3, 527-549.

Barton, J.J.S., Zhao, J.H., & Keenan, J.P. (2003). Perception of global facial geometry in the inversion effect and prosopagnosia. Neuropsychologia, 41 (12), 1703-1711.

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

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A.M., & Miller, P. (1999). Verification of face identities from images captured on video. Journal of Experimental Psychology: Applied, 5, 339-360.

Brunelli, R., & Poggio, T. (1993). Face Recognition - Features Versus Templates. Ieee Transactions on Pattern Analysis and Machine Intelligence, 15 (10), 1042-1052.

Burbeck, C.A. (1991). Encoding spatial relations. In: R.J. Watt (Ed.) Pattern Recognition by Man and Machine, 14, London: Macmillan.

Burton, A.M., Jenkins, R., Hancock, P.J.B., & White, D. (2005). Robust representations for face recognition: The power of averages. Cognitive Psychology, 51 (3), 256-284.

Burton, A.M., Miller, P., Bruce, V., Hancock, P.J.B., & Henderson, Z. (2001). Human and automatic face recognition: a comparison across image formats. Vision Research, 41 (24), 3185-3195.

Cabeza, R., & Kato, T. (2000). Features are also important: contributions of featural and configural processing to face recognition. Psychological Science, 11 (5), 429-433.

Collishaw, S.M., & Hole, G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. Perception, 29 (8), 893-909.

Diamond, R., & Carey, S. (1986). Why Faces Are and Are Not Special - an Effect of Expertise. Journal of Experimental Psychology-General, 115 (2), 107-117.

Dodgson, N.A. (2004). Variation and extrema of human interpupillary distance. In: A.J. Woods, J.O. Merritt, S.A. Benton, & M.T. Bolas (Eds.), Stereoscopic Displays and Virtual Reality Systems XI, 5291 (pp. 36-46). San Jose: Proceedings of SPIE.

Farkas, L.G. (1981). Anthropometry of the Head and Face in Medicine. New York: Elsevier.

Freire, A., & Lee, K. (2001). Face recognition in 4- to 7-year-olds: processing of configural, featural, and paraphernalia information. Journal of Experimental Psychology-General, 115 (2), 107-117.

Freire, A., Lee, K., & Symons, L.A. (2000). The face-inversion effect as a deficit in the encoding of configural information: direct evidence. Perception, 29 (2), 159-170.

Gauthier, I., Williams, P., Tarr, M.J., & Tanaka, J. (1998). Training 'greeble' experts: a framework for studying expert object recognition processes. Vision Research, 38 (15-16), 2401-2428.

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

Graham, C.H. (1965). Vision and Visual Perception. New York: John Wiley & Sons.

Haig, N.D. (1984). The effect of feature displacement on face recognition. Perception, 13, 505-512.

Hess, R.F., Barnes, G., Dumoulin, S.O., & Dakin, S.C. (2003). How many positions can we perceptually encode, one or many? Vision Research, 43 (14), 1575-1587.

Jeffery, L., Rhodes, G., & Busey, T. (2006). View-specific coding of face shape. Psychological Science, 17 (6), 501-505.

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. Perception, 19 (4), 531-543.

Le Grand, R., Mondloch, C.J., Maurer, D., & Brent, H.P. (2001). Neuroperception. Early visual experience and face processing. Nature, 410 (6831), 890.

Le Grand, R., Mondloch, C.J., Maurer, D., & Brent, H.P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. Nature Neuroscience, 6 (10), 1108-1112.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. Quarterly Journal of Experimental Psychology Section: Human Experimental Psychology, 53 (2), 513-536.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. Perception, 30 (1), 73-83.

Liu, C.H., & Chaudhuri, A. (2003). Face recognition with perspective transformation. Vision Research, 43 (23), 2393-2402.

Martin, D.K., & Holden, B.a. (1982). A New Method for Measuring the Diameter of the Invivo Human Cornea. American Journal of Optometry and Physiological Optics, 59 (5), 436-441.

Maurer, D., Le Grand, R., & Mondloch, C.J. (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6 (6), 255-260.

Mondloch, C.J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. Perception, 31 (5), 553-566.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. Vision Research, 39 (23), 3824-3833.

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

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. Perception, 17 (1), 43-63.

Searcy, J.H., & Bartlett, J.C. (1996). Inversion and processing of component and spatial-relational information in faces. Journal of Experimental Psychology: Human Perception and Performance, 22 (4), 904-915.

Sekuler, A.B., Gaspar, C.M., Gold, J.M., & Bennett, P.J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. Current Biology, 14 (5), 391-396.

Steyvers, M., & Busey, T.A. (2001). Predicting Similarity Ratings to Faces Using Physical Descriptions. In M.J. Wenger, & J.T. Townsend (Eds.), Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges Hillsdale, NJ: Lawrence Erlbaum Associates.

Theodorff, C.D., & Lowther, G.E. (1990). Quantitative effect of optic zone diameter changes on rigid gas permeable lens movement and centration. ICLC, 17, 92-95.

Tyler, C.W., & Chen, C.-C. (2006). Spatial summation of face information. Journal of Vision, 6 (10), 1117-1125.

Young, J.W. (1993). Head and Face Anthropometry of Adult U.S. Civilians. U.S. Department of Transportation, Federal Aviation Administration.

Yovel, G., & Duchaine, B. (2006). Specialized face perception mechanisms extract both part and spacing information: evidence from developmental prosopagnosia. Journal of Cognitive Neuroscience, 18 (4), 580-593.

Zhao, W., Chellappa, R., Phillips, P.J., & Rosenfeld, A. (2003). Face recognition: A literature survey. ACM Computing Surveys, 35 (4), 399-459.

Chapter 5

Summary & Future Directions

5.1 Summary

Chapters 2 and 3 of this thesis were a direct follow-up to the classification image study of Sekuler et al. (2004). Sekuler et al. demonstrated that observers use similar regions around the eyes and eyebrows to identify both upright and inverted faces. Their results are also consistent with the idea that observers rely strongly on a template-matching strategy to judge the appearance of both upright and inverted faces. Based on these findings, and a quantitative analysis of their classification images, Sekuler et al. speculated that inverted face identification is less accurate than upright face identification because of a subtle, but systematic misuse of information around the eyes and eyebrows. The results of Chapter 2 provide strong support for this conclusion. Chapter 3 examined the possibility that perhaps more dramatic differences between upright and inverted face processing could be found in the spatial frequency domain, as opposed to the how different regions of the face are used. However, the results presented in Chapter 3 demonstrate that upright and inverted face identification rely on the same narrow band of spatial frequencies center on approximately 9 cycles per face width. In order to take account of this result, any explanation of the face inversion effect must consider how inverted face identification might misuse information within the same narrow band of spatial frequencies that are used to identify upright faces. Therefore, the results of the Chapter 3, together with those of the Chapter 2, provide strong evidence against the prevailing notion that upright and inverted face recognition rely on qualitatively different processes. As was explained in the Chapter 1, the idea that upright and inverted face identification are qualitatively different is a central assumption underlying so-called configural or relational theories of face recognition. Therefore, the findings presented in Chapters 2 and 3 also call into question the idea that relational cues (variations in the exact spacing among facial features) are critical for normal face recognition. In order to directly assess the importance of relational cues, the Chapter 4 examined if human sensitivity to such cues is fine enough to resolve the limited variation in relational cues that are found in real faces. The results of the experiments described in Chapter 4 suggest that two of the most often studied relational cues, inter-pupillary distance and nose-to-mouth-distance, would not be particularly useful for face identification in natural contexts. Moreover, we found evidence suggesting that the skills underlying judgments of facial feature-spacing do not contribute significantly to accuracy in a realistic face identification task.

5.2 Applications to the Thatcher Illusion

The face inversion effect has been primarily used as an experimental tool for determining those elements of facial information that are most important for face identification. However, this approach rests on the assumption that upright and inverted face identification are qualitatively different. Chapters 2 and 3 present evidence that is contrary to this assumption, and thus call into question the use of the face inversion effect as a tool for uncovering critical facial characteristics. However, while our results question the reason why the face inversion effect is normally studied, our conclusions do not make the face inversion effect any less interesting. The work presented in Chapters 2 and 3, together with the classification image results of Sekuler et al., demonstrate that upright and inverted face identification, despite being very different in accuracy, rely on the very similar regions of the face, and also the same approximate band of spatial frequencies. Therefore, our findings suggest that rather subtle changes in the use of facial information can have strong and reliable consequences on identification accuracy. However, face inversion can do much more than just reduce our ability to identify faces. The Thatcher Illusion demonstrates that face inversion can also drastically reduce our ability to perceive distortions of facial images that are normally perceived as highly grotesque. To experience the strong phenomenology of the Thatcher Illusion, please see the facial photo shown in the middle of Figure 5.1. The eyes and the mouth in this face have been inverted while the rest of the face was left in its normal orientation. For historical reasons, this method of distorting the face has been called 'Thatcherization', and the end product is sometimes called a 'Thatcher' face. The distortions introduced by Thatcherization are very clear and frequently evoke a strong percept of grotesqueness. Of course, there is the possibility that we chose a face that was very grotesque to begin with, and would remain grotesque no matter how we rearranged its facial features. To make it clear that this is not the case, please see the original facial image in Figure 5.1, which is the first one from the left. It should be obvious that the normal face and its Thatcherized version are very different. However, once both images are turned upside-down, it is no longer immediately clear that one of these inverted faces has been distorted. After a close inspection of the face, one will eventually be able to identify the distortion that was made but, nevertheless, the strong percept of grotesqueness cannot be recovered after face inversion. The inability to experience the grotesque percept of a Thatcherized face after it has been inverted is referred to as the Thatcher Illusion. The Thatcher Illusion is perhaps one of the most widely cited experimental findings that face researchers use to support the idea that upright and inverted face processing are qualitatively different. The findings of this thesis demonstrate that very subtle changes in facial information accompany face inversion and that these changes are sufficient to explain the effect of inversion on identification accuracy. However, the Thatcher Illusion demonstrates that face inversion can virtually eliminate our ability to experience a strong percept of grotequeness. Can the effect of inversion on the Thatcher Illusion also be explained by very subtle changes in the use of facial information, rather than the gross changes in information suggested by prevailing theories? I demonstrate that this is indeed possible.

The image on the far right of Figure 5.1 is an image of the difference between the normal and Thatcherized faces shown in the same figure. In other words, the first face was superimposed on the second, and a pixel-by-pixel subtraction was performed in order to obtain this difference image. The difference image demonstrates that it is not necessary to look at the entire face, or to make judgments about the spacing among facial features, in order to tell a normal face apart from its Thatcherized version. In fact, the difference between a normal face and its Thatcherized version resides solely within a small region of both eyes, and a small region encompassing the mouth. It must be clear that I am not making any claims about where in the face the grotesque percept associated with Thatcherization originates. However, I am asserting that it is necessary to perceive the changes in local features that define the physical effect of Thatcherization before one can even have a percept of grotesqueness. In other words, one must detect Thatcherization before one can experience the grotesqueness of Thatcherization. Therefore, it is entirely possible that an inability to perceive subtle changes within the eyes and mouth regions will drastically reduce the percept of grotesqueness in a Thatcherized face. The results of Chapter 1 of this thesis, together with the classification image results of Sekuler et al. (2004), demonstrate that face inversion is associated with very subtle changes in how observers use information within local regions around the eyes and eyebrows. These quantitative changes in face processing are sufficient to explain the reduction in identification efficiency that accompanies face inversion. One implication of these findings is that observers should find it difficult to discriminate individual facial features when they are shown in an inverted face, especially the eyes. The task of detecting a Thatcherized face is similar to that of discriminating different facial features, except that the features differ in their local orientation. According to the results of this thesis,

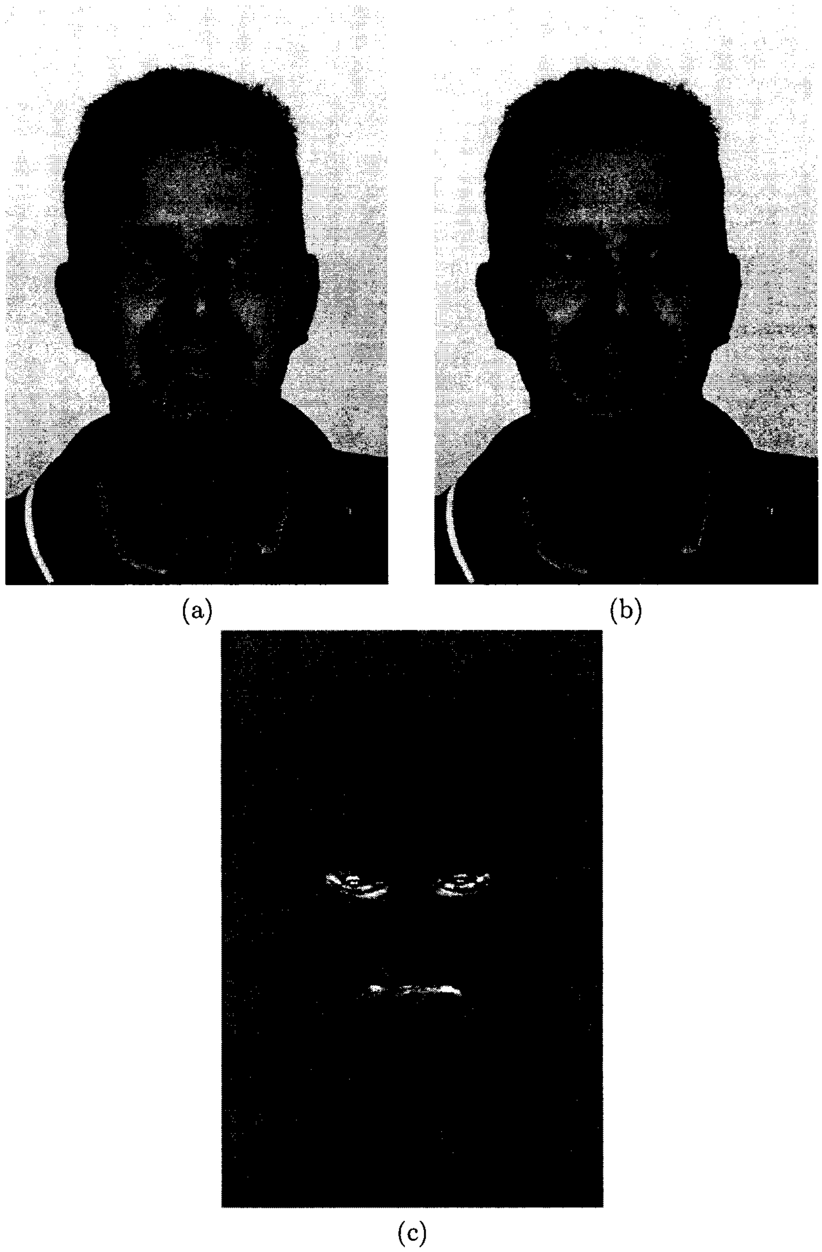


Figure 5.1: A normal face (a), its Thatcherized version (b), and the image of the Thatcherized face subtracted from the normal face (c).

one might predict that observers will find it difficult to discriminate Thatcherized features from normal features when they are shown in an inverted face. It is important to note that this is not an explanation of the inversion effect. If an observer is devoting all their attention to just one eye, the task of discriminating a Thatcher face from its normal version is exactly the same whether or not both faces are shown upright, or inverted. Therefore, observers must be influenced by more than just those local features that have been Thatcherized. Nonetheless, an observer's ability to discriminate changes within those local features is critical for obtaining the percept of grotesqueness.

My description of how inversion alters the information-processing of Thatcherized faces has rested on the assumption that the processes underlying face identification are similar to those underlying the detection of Thatcherization. If this assumption is true, then observers should be using similar spatial frequency components to identify faces, and also to detect Thatcher faces. The results presented in Chapter 3 demonstrate that both upright and inverted face identification rely on the same narrow band of spatial frequencies in the face, and that this 2-octave wide band is centered at around 9 cycles per face width; previous studies have measured the spatial frequency tuning of upright face identification and found similar results (Gold, Bennett & Sekuler, 1999, Näsänen, 1999). No one has ever measured the spatial-frequency tuning of Thatcher face detection. However, I can demonstrate that successful detection of Thatcherized eyes need only utilize a narrow band of spatial frequencies, and that this 2-octave wide band is centered at around 9 cycles per face width. As mentioned before, the discrimination of a Thatcher face from its normal version requires that observers use the information contained in the difference image shown on the left side of Figure 5.1. Therefore, the 2D Fourier transform of this difference image tells us how the relevent information is distributed across the spatial-frequency spectrum. The average spectra of the Thatcher information contained in the eyes of 14 faces is shown in Figure 5.2, along with standard error bars. While there is variation in the exact amount of power distributed across frequency, one can see a very strong central ten-

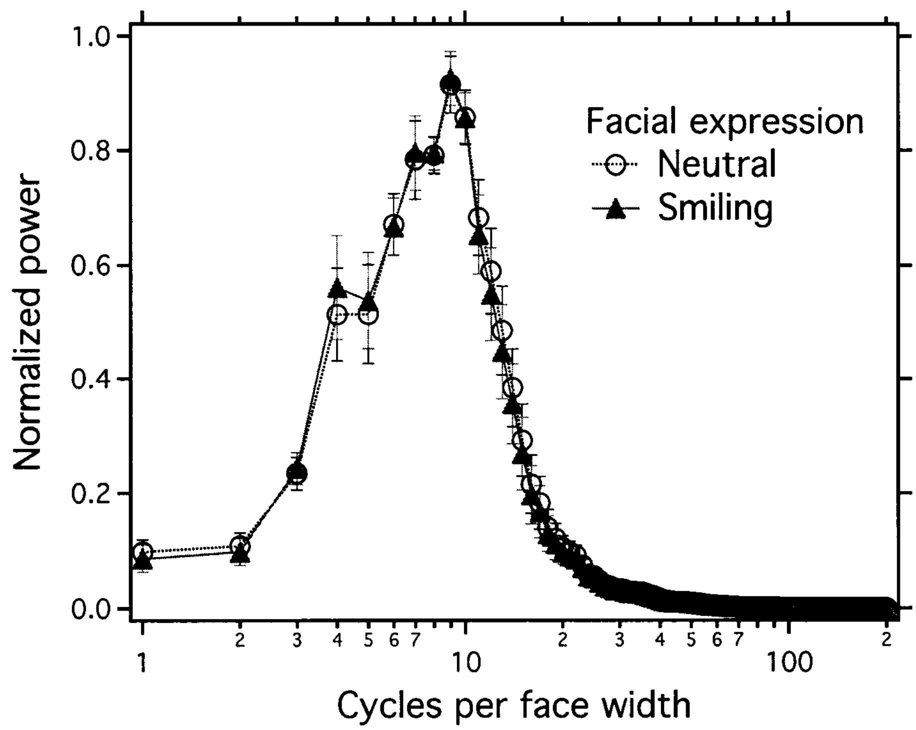


Figure 5.2: Normalized power spectra of eyes-only Thatcher information for 14 faces. Thatcher information is the image formed by the subtraction of the eyes-only Thatcher face from its unedited version (similar to that shown in Figure 5.1, except with no information in the mouth region). Hollow circles depict the mean spectrum ($\pm$ SE) across 7 faces with neutral expressions, and filled triangles depict mean spectrum ($\pm$ SE) across the same 7 faces but smiling. Thatcher information is narrowband for both neutral and smiling faces, peaking at 9 cycles per face width, and less than 2 octaves wide. Face width is measured by the bizygion diameter, which is the widest part of the visible face (Farkas, 1981).

dency at around 9 cycles per face width, with power falling off rapidly above and below this band. This bandpass property of Thatcher eye information demonstrates that the information processing strategy observers should be using to optimally detect Thatcherized eyes is the same strategy that observers do employ in order to identify faces. Future experiments will need to measure the spatial frequency tuning of Thatcher eye detection in order to confirm this hypothesis.

5.3 Conclusion

The idea that subtle, quantitative changes in information processing can have profound effects on both the discriminability and the phenomenology of faces has the potential to synthesize various results in the face identification literature. For example, the relationship that we have just observed between the information contained in Thatcherized eyes and the bandpass nature of its power spectrum suggest that the exact tuning of spatial frequency may be related in a systematic way to local features of the face. For both Thatcher detection and face identification, there appears to be a relationship between the informative structure of the eyes and a band of spatial frequencies around 9 cycles per face width. This observation is significant because currently there are no explicit theories for why face identification relies on the particular band of frequencies that it does. The classification image results of Sekuler et al. (2004) can be interpreted to mean that the use of eyes and eyebrows leads to bandpass tuning. However, our current analysis of the Thatcher information spectra suggest that this is a consistent relationship between eye structure and bandpass tuning, and they also help us to visualize the relevant structures in the eye.

5.4 References

Barton, J.J., Keenan, J.P., & Bass, T. (2001). Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. Br

J Psychol, 92 Part 3, 527-549.

Barton, J.J.S., Zhao, J.H., & Keenan, J.P. (2003). Perception of global facial geometry in the inversion effect and prosopagnosia. Neuropsychologia, 41 (12), 1703-1711.

Collishaw, S.M., & Hole, G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. Perception, 29 (8), 893-909.

Diamond, R., & Carey, S. (1986). Why Faces Are and Are Not Special - an Effect of Expertise. Journal of Experimental Psychology-General, 115 (2), 107-117.

Farah, M.J., Tanaka, J.W., & Drain, H.M. (1995a). What Causes the Face Inversion Effect. Journal of Experimental Psychology-Human Perception and Performance, 21 (3), 628-634.

Farah, M.J., Wilson, K.D., Drain, H.M., & Tanaka, J.R. (1995b). The Inverted Face Inversion Effect in Prosopagnosia - Evidence for Mandatory, Face-Specific Perceptual Mechanisms. Vision Research, 35 (14), 2089-2093.

Gauthier, I., Williams, P., Tarr, M.J., & Tanaka, J. (1998). Training 'greeble' experts: a framework for studying expert object recognition processes. Vision Res, 38 (15-16), 2401-2428.

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

Kanwisher, N. (2006). Neuroscience. What's in a face? Science, 311 (5761), 617-618.

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. Perception, 19 (4), 531-543.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 53 (2), 513-536.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. Perception, 30 (1), 73-83.

Martelli, M., Majaj, N.J., & Pelli, D.G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. Journal of Vision, 5 (1), 58-70.

Maurer, D., Le Grand, R., & Mondloch, C.J. (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6 (6), 255-260.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. Vision Res, 39 (23), 3824-3833.

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. Perception, 17 (1), 43-63.

Searcy, J.H., & Bartlett, J.C. (1996). Inversion and processing of component and spatial-relational information in faces. J Exp Psychol Hum Percept Perform, 22 (4), 904-915.

Thompson, P. (1980). Margaret Thatcher: a new illusion. Perception, 9 (4), 483-484.

Yin, R.K. (1969). Looking at upside-down faces. Journal of Experimental Psychology, 81, 141-145.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configurational information in face perception. Perception, 16 (6), 747-759.