

Spatial Frequencies Underlying Upright and Inverted Face Identification

by

Verena Willenbockel
B.Sc. University of Osnabrück, 2006

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCES

in the Department of Psychology

© Verena Willenbockel, 2008
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy
or other means, without the permission of the author.

Spatial Frequencies Underlying Upright and Inverted Face Identification

by

Verena Willenbockel
B.Sc. University of Osnabrück, 2006

Supervisory Committee

Dr. James Tanaka, Supervisor
Department of Psychology

Dr. Clay Holroyd, Departmental Member
Department of Psychology

Dr. Stephen Lindsay, Departmental Member
Department of Psychology

Dr. Quoc Vuong, Outside Member
Institute of Neuroscience, Newcastle University

Supervisory Committee

Dr. James Tanaka, Supervisor
Department of Psychology

Dr. Clay Holroyd, Departmental Member
Department of Psychology

Dr. Stephen Lindsay, Departmental Member
Department of Psychology

Dr. Quoc Vuong, Outside Member
Institute of Neuroscience, Newcastle University

Abstract

The face inversion effect (FIE; Yin, 1969) raises the question of whether upright face identification is mediated by a special mechanism that is disrupted by inversion. The present study investigates the effect of face inversion on the perceptual encoding of spatial frequency (SF) information using a novel variant of the *Bubbles* technique (Gosselin & Schyns, 2001). In Experiment 1, the *SF Bubbles* technique was validated using a simple plaid detection task. In Experiment 2, SF tuning of upright and inverted face identification was measured. While the data showed a clear FIE (28% higher accuracy and 455 ms shorter reaction times for upright faces), SF tunings were remarkably similar in both conditions ($r = .96$; a single SF band of ~ 2 octaves peaking at ~ 9 cycles per face width). Experiments 3 and 4 demonstrated that *SF Bubbles* is sensitive to bottom-up and top-down induced changes in SF tuning, respectively. Overall, the results show that the same SFs are utilized in upright and inverted face identification, albeit not with equal efficiency.

Table of Contents

Supervisory page	ii
Abstract	iii
Table of Contents	iv
List of Figures	v
Acknowledgments	vi
Introduction	1
Experiment 1	10
Experiment 2	16
Experiment 3	27
Experiment 4	30
General Discussion	41
Conclusion	46
References	47

List of Figures

Figure 1	8
Figure 2	12
Figure 3	17
Figure 4	19
Figure 5	21
Figure 6	23
Figure 7	25
Figure 8	26
Figure 9	31
Figure 10	33
Figure 11	36
Figure 12	38
Figure 13	39

Acknowledgments

I would like to thank my supervisor J. Tanaka, as well as D. Fiset and F. Gosselin for helpful discussion and for comments on a draft of this thesis. This research was funded by a scholarship from the German Academic Exchange Service (DAAD) to the author and grants from the Natural Sciences and Engineering Research Council of Canada, the National Science Foundation, and the James S. McDonnell Foundation to J. Tanaka.

Introduction

Humans typically recognize each other by looking at the face (e.g., Ruiz-Soler & Beltran, 2006). Adults can recognize thousands of individuals by their faces rapidly and effortlessly, even under poor lighting conditions and from a wide range of viewpoints (e.g., Maurer, Le Grand, & Mondloch, 2002). An exception is turning faces upside-down: recognition accuracy decreases considerably and response latencies increase when faces are rotated by 180° in the picture-plane (e.g., Yin, 1969; Diamond & Carey, 1986). This drop in recognition performance is believed to be larger for faces than for other mono-oriented objects (e.g., houses and airplanes) in untrained individuals, and is commonly called the face inversion effect (FIE; Yin, 1969).¹

Gaining insights into the factors that enable us to recognize a face has motivated both basic and applied research, for example, to better understand the interaction between visual information from the outside world and recognition processes in the brain, to design efficient video surveillance systems, and improve the work with eyewitnesses (Loftus & Harley, 2004; Ruiz-Soler & Beltran, 2006). The FIE has received much interest because it is thought to provide a window into the mechanisms underlying effective face processing. Previous research has suggested that upright face recognition engages a specific mechanism which is disrupted or inhibited by inversion (reviews in Rossion, in press; Rossion & Gauthier, 2002; Valentine, 1988). Since upright and inverted faces are of the same complexity, and almost identical in their low-level properties, such as luminance, contrast, and spatial frequencies (only phase information differs), the drop in recognition performance with inversion cannot easily be attributed to

¹ For a comparably large inversion effect with faces and human body positions see Reed, Stone, Bozova, & Tanaka (2003).

stimulus properties *per se*. It has thus been of interest to examine exactly how the processing of upright and inverted faces differs and why it differs more than with other objects.

The FIE is one of the most robust phenomena reported in the face processing literature (Rossion, in press). It can be observed for both unfamiliar and familiar faces (Collishaw & Hole, 2000) and in a variety of experimental conditions. For instance, it has been demonstrated in experiments using blocked presentation (Valentine & Bruce, 1986a) and randomized presentation (Scapinello & Yarmey, 1970; Yin, 1969), in old-new recognition tasks (e.g., Carey, Diamond, & Woods, 1980; Scapinello & Yarmey, 1970) as well as in two-alternative forced-choice paradigms with or without delay (e.g., Freire, Lee, & Symons, 2000; Leder & Bruce, 2000; Tanaka & Farah, 1993; Yin, 1969). However, even after three decades of research, there is no consensus about how to explain all the results obtained (e.g., Ruiz-Soler & Beltran, 2006).

Numerous studies have found that inversion results in particular difficulties for the processing of the metric distance between facial features (e.g., inter-ocular distance, mouth-nose distance) which define configural information of faces. For instance, discrimination (i.e., same/different) performance is adversely affected by inversion to a greater degree when “different” stimulus pairs vary only on configural information than when they differ on the shape or color of the facial features (featural information) (e.g., Freire, Lee, & Symons, 2000; Leder et al., 2001; Leder & Bruce, 1998). A recent study indicated that inversion also impairs the perception of the featural information contained in the lower part of the face (Tanaka, Kaiser, Bub, & Pierce, in preparation). The nature of the processing difficulties caused by face inversion has been suggested to result from a

disruption of holistic processing which is defined as “the simultaneous integration of the multiple features of a face into a single perceptual representation” (Rossion, in press, p. 5). This view is congruent with the part-whole recognition effect, that is, the higher recognition accuracy of a facial feature when shown in the context of a whole face than when presented in isolation (Tanaka & Farah, 1993; Tanaka & Sengco, 1997). It is also congruent with the face composite effect—the difficulty to identify the top or bottom half of a face when joined with the complementary half of another face (e.g., Hole 1994; Young, Hellawell, & Hay, 1987). The composite effect was found to be equally strong for faces shown at angles of rotation between 0° and 60°, and to be reduced dramatically at angles from 90° to 180° (Rossion & Boremanse, in press). Overall, these findings seem to suggest that inversion differentially affects certain processes during face perception. Specifically, the notion of qualitative processing differences is that upright faces are processed holistically, whereas inverted faces are processed on the basis of individual features (reviews in Rossion, in press; Rossion & Gauthier, 2002; Valentine, 1988).

Alternatively, it has been claimed that the processing differences underlying the FIE are quantitative rather than qualitative. Sekuler et al. (2004) employed a response classification technique to investigate the perceptual strategies used for upright and inverted face discrimination. In response classification, Gaussian white noise is added on each trial to the stimuli that observers are asked to classify. After a sufficient number of trials, a map showing the linear association between each pixel’s contrast and the observer’s responses can be constructed (a “classification image”). Sekuler et al.’s classification images showed that observers rely to a great extent on eyes and eyebrows when identifying both upright and inverted faces (see also Gold, Sekuler, & Bennett,

2004; Gosselin & Schyns, 2001; Gosselin & Schyns, 2005; Schyns, Bonnar, & Gosselin, 2002; Williams & Henderson, 2007). Their results also revealed that observers' performance was slightly more efficient than what could be inferred from their classification images alone, suggesting that observers used information not captured by the linear relationship between each pixel's contrast and the observers' responses. However, the estimated contribution of these nonlinearities (for details see Murray, Bennett, & Sekuler, 2005) was also similar in the upright and inverted condition. At the same time, a clear FIE was present. Observers needed more contrast (i.e., information surviving the noise) in inverted faces to maintain accuracy at the same level than in upright faces. Processing efficiency, calculated by the normalized cross-correlation between the classification images of an ideal discriminator and human observers, was lower for inverted faces. Sekuler et al. (2004) concluded that inversion did not result in the use of qualitatively different processing modes, but instead, in a reduced efficiency for the extraction of the same cues.

Besides the reliance on the eye region, upright face identification also relies on a narrow range of spatial frequencies (SFs) (see Morrison & Schyns, 2001, for a review). The frequency with which light-dark transitions repeat across an image can be measured in cycles per stimulus (e.g., cycles per image (cpi), cycles per face (cpf)) or in cycles per degree of visual angle that the stimulus spans on the retina. High SFs represent the fine-grained information in the stimulus such as the eyelashes or edges of the mouth. In contrast, low SFs convey coarse information such as luminance blobs and blurred shapes. In general, the human visual system analyzes the complex luminance variations that make up the visual stimulus with discrete channels, each tuned to a specific SF range (see

De Valois & De Valois, 1990, for a review). Several studies have shown that the effective recognition of upright faces relies on a band of SFs between 8 and 16 cpf (e.g., Costen, Parker, & Craw, 1994, 1996; Gold, Sekuler, & Bennett, 1999b; Näsänen, 1999; reviews in Morrison & Schyns, 2001; Ruiz-Soler & Beltran, 2006), whereas object discrimination, even at the individual level, was found to be based on a broader SF band (Biederman & Kalocsai, 1997). The hypotheses addressed by Sekuler et al. (2004) may thus be extended from the spatial to the SF domain: Are observers worse at identifying inverted than upright faces because they rely on different SFs in the two conditions (qualitative difference hypothesis) or because they are less efficient at processing the same SF information in the inverted than the upright condition (quantitative difference hypothesis)?

Indirect support for the qualitative difference hypothesis in the SF domain comes from recent studies investigating the role of different SF ranges in featural compared to configural and holistic processing of faces. Goffaux et al. (2005) demonstrated that participants performed better in a matching task with faces differing in featural information when using high-pass (above 32 cpf) than low-pass filtered faces (below 8 cpf). In contrast, for faces that differed in configural information, performance was better for low-pass than high-pass filtered stimuli. Similarly, a study by Goffaux and Rossion (2006) based on the part-whole recognition effect and the composite effect which are thought to reflect holistic processing, found that these effects were largely supported by low SFs. One might thus expect to find a larger reliance on low SFs for upright faces, based on the assumption that they are processed largely holistically, than for inverted faces which are suggested to be processed mainly by features.

In contrast, the view defended by Sekuler et al. (2004) is that the processing differences between upright and inverted faces are primarily quantitative. Specifically, they argue that the same facial information is used for upright and inverted faces, but that processing efficiency is lower in the inverted condition. A corollary of this position is that the recognition of upright and inverted face recognition should be mediated by the same SF channel. The response classification data of Sekuler et al. (2004) are insufficient to determine whether the same SFs are used in upright and inverted face recognition. Theoretically, it would be possible for the eye and eyebrow regions revealed in their upright and inverted classification images to hide different patterns of SF use (e.g., 8 cpf for upright eye and eyebrow regions and 16 cpf for those same regions but inverted). Thus, investigating the effect of inversion on SF use could provide important insights into the nature of the processing differences underlying the FIE.

Few studies have addressed directly which role SFs play in the FIE. An unpublished study by Nakayama (2003) based on a narrow band masking noise approach provided support for a change in SF information use with inversion. Their results of a 4-alternative discrimination task indicated that the processing of inverted faces relied on a broader range of SFs than upright faces. In contrast, an unpublished study by Gaspar, Sekuler, and Bennett (2005) based on critical band masking showed that observers used a similar narrow SF band centered at approximately 8 cpf for upright and inverted faces in a 10-alternative identification task, while significantly more contrast was needed to identify inverted faces. Thus support for both orientation-dependency and -independency of SF tuning has been reported, and more research is needed to unravel the effect of face inversion on SF use.

To test whether inversion leads to changes in SF tuning, we used a novel SF variant of the Bubbles technique (Fiset, Blais, Gosselin, & Schyns, 2006; Gosselin & Schyns, 2001; see McCotter et al., 2005, for a distinct attempt at applying Bubbles to SFs in natural scenes). The Bubbles technique has been used in the spatial domain to reveal the location of the visual cues that are diagnostic in particular categorization tasks (e.g., Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002). The method consists of sampling the stimulus information on a trial-by-trial basis using a number of randomly located Gaussian windows (“bubbles”). The bubbles reveal a portion of the image to an observer who may use this information for performing the categorization task. After a sufficient number of trials, multiple linear regressions performed on the bubbles’ location and response accuracy, response time or any other measure of interest reveal the effective stimulus. This technique has been applied to full-spectrum images (e.g., Gosselin & Schyns, 2001, Experiment 1) as well as band-pass filtered images (e.g., Gosselin & Schyns, 2001, Experiment 2; Schyns, Bonnar, & Gosselin, 2002; see Figure 1a for an example). However, it has not been used previously for sampling the SF content of images.

In this study, we applied the Bubbles technique in the SF domain. Analogous to randomly varying the availability of local cues in the image, SF Bubbles randomly varies the availability of SFs on a trial-by-trial basis (Figure 1b). The main strength of the SF Bubbles technique is that it minimizes the risk that participants adapt to a predictable stimulus manipulation such as high-pass, low-pass, or band-pass filtering (e.g., Goffaux et al., 2005) or critical band noise masking (e.g., Gaspar, Sekuler, & Bennett, 2005; Solomon & Pelli, 1994). Furthermore, it does not assume that observers only process one



Figure 1. Illustration of the Bubbles technique applied in the spatial domain (a) and the SF domain (b; *SF Bubbles*) on three hypothetical trials. In the spatial domain, the face is partly revealed by a mid-grey mask with a number of randomly located Gaussian windows. In the SF domain, all local parts of the face are visible but the SF composition of the face is randomly sampled on each trial. For details see the *Spatial Frequency Bubbles* section of Experiment 1.

SF band at a time, which also distinguishes the SF Bubbles technique from other methods employed to assess the use of SF information (e.g., Gaspar, Sekuler, & Bennett, 2005; Solomon & Pelli, 1994).

The present thesis comprises three experiments. The first experiment was designed to assess the validity of the SF Bubbles technique. Using a plaid (i.e., the sum of two sine wave gratings) detection task, it was verified that the SF Bubbles method can uncover precisely the SFs of the plaid. The second experiment, which is the main experiment, examined which SFs are diagnostic for the accurate (Experiment 2a) and fast (Experiment 2b) identification of 20 grayscale face photos (10 identities x 2 exemplars) presented upright or inverted. The goal of this experiment was to reconstruct the SF filters used by human observers to effectively perform the face identification task and thus test the quantitative vs. qualitative difference hypothesis at the SF level. To anticipate the main result, no difference in SF use was revealed between the upright and inverted condition. In order to rule out that this null result was due to an insensitivity of the SF Bubbles technique to bottom-up or top-down influences on SF tuning, two follow-up experiments were carried out. Specifically, Experiment 3 re-examined SF tuning in the identification task of Experiment 2 as a function of retinal image size, which is known to influence SF tuning in a bottom-up fashion (e.g., Majaj et al., 2002). In Experiment 4, task demands were modified (gender or happy vs. neutral discrimination) to modulate SF tuning in a top-down fashion (e.g., Schyns & Oliva, 1999). In both cases, subtle differences in SF use were revealed, confirming that the SF Bubbles technique is sensitive to bottom-up and top-down induced changes in SF tuning.

In sum, the main goals of the current study were (1) to introduce a novel technique that reveals the use of SFs and is applicable in a variety of contexts, including face processing, letter and word recognition, as well as object and scene categorization, and (2) to use this technique to provide insights into the perceptual encoding stage of face recognition.

Experiment 1

The Bubbles technique was introduced as “a general technique that can assign the credit of a categorization performance to specific visual information” (Gosselin & Schyns, 2001, p. 2261). Here, we transposed the approach in the SF domain. Specifically, the SFs that make up a given visual stimulus served as the search space for diagnostic information.

The purpose of the first experiment was to test whether the technique can reveal precisely the SFs that convey the information diagnostic for the task. For that purpose, we employed a simple plaid (i.e., the sum of two SFs) detection task. If the SF Bubbles technique works adequately, then it should be possible to recover the SFs comprised in the plaid.

Method

All the experiments reported in this thesis were run on a dual core 2.93 GHz PC using a program written in MatLab 7.4 (Natick, MA, USA) with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a 22-inch Viewsonic CRT monitor at a refresh rate of 85 Hz. The resolution was set to 1024 x 768 pixels (except in Experiment 3). The monitor was calibrated to allow a linear manipulation of luminance. The resulting corrected table contained 154 luminance levels, ranging from

0.3 cd/m² to 98.7 cd/m². The background luminance was 49.3 cd/m² (corresponding to mid-gray).

Participants were one male and two female University of Victoria students aged between 22 and 25 years ($M = 24$ years). All participants had normal or corrected-to-normal vision, and two were naïve as to the purpose of the experiment, whereas the third was the author (Observer 1). Participants gave informed consent approved by the University of Victoria Human Research Ethics Committee.

Participants were instructed to perform a plaid detection task. The original plaid was comprised of a horizontal sine wave grating with a SF of 10 cycles per image (cpi) and a vertical sine wave grating with a SF of 45 cpi (Figure 2) and had a size of 256 x 256 pixels. On “signal present” trials (50% of trials), the plaid was sampled with the SF Bubbles technique (see below) and displayed embedded in Gaussian white noise. On “signal absent” trials, a Gaussian white noise field of 256 x 256 pixels was displayed. Each trial began with a central fixation cross, followed by the stimulus presented for 870 ms, and then by a homogeneous mid-gray field that remained on the screen until the observer responded by pressing the appropriate key on a computer keyboard. “Signal present” and “signal absent” trials occurred in random order. No feedback was provided. Each observer performed eleven 100-trial blocks in a one-hour session with breaks between blocks. Participants were seated in a dark room and a chin rest was used to maintain viewing distance at 53 cm—stimuli subtended a visual angle of 10.24° x 10.24°.

Spatial Frequency Bubbles

All experiments reported in this thesis revealed SF use by employing the SF Bubbles technique. This section describes SF Bubbles in general and illustrates the use

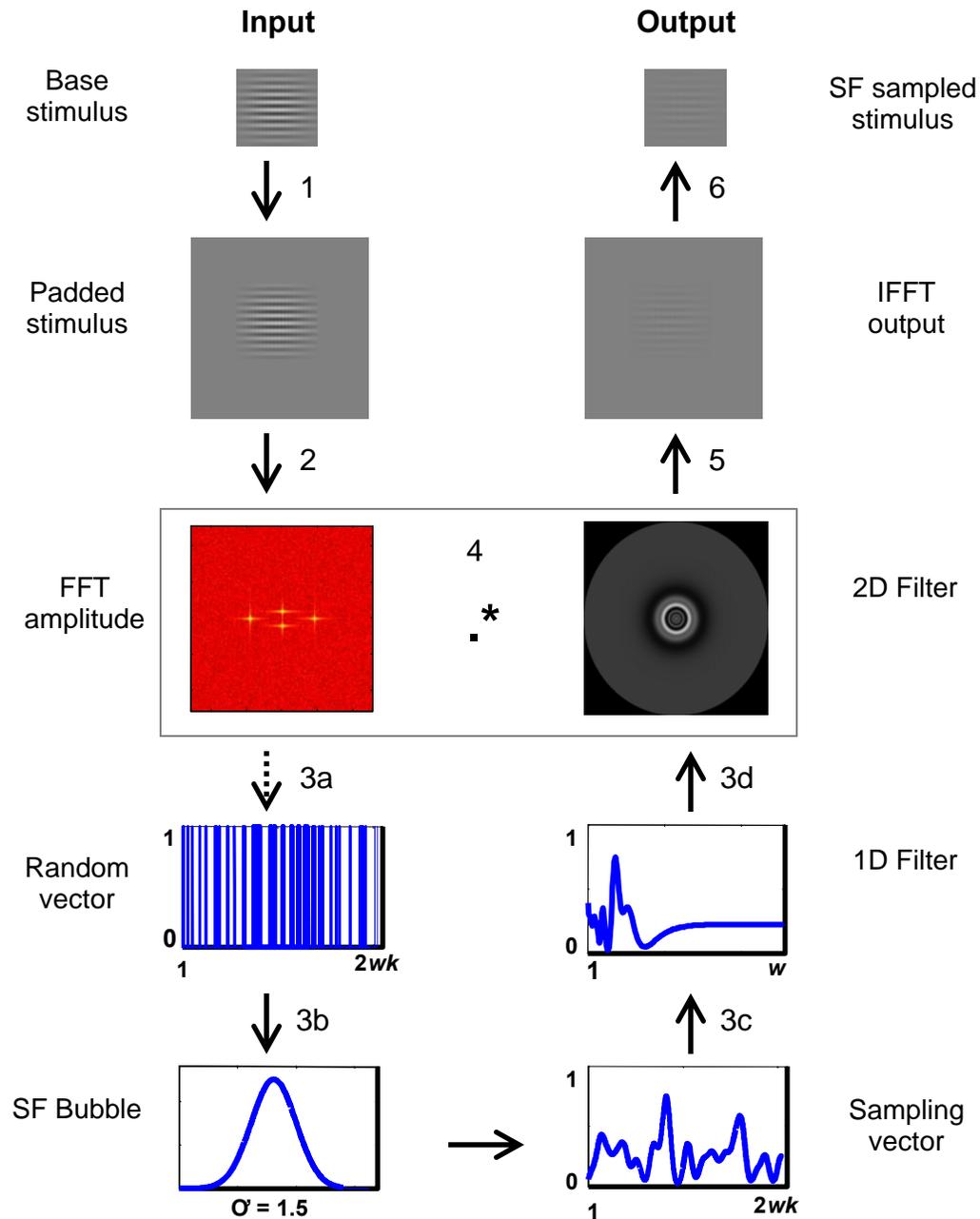


Figure 2. Illustration of the SF Bubbles technique. (1) Padding. (2) Fast Fourier Transform. (3) Construction of random SF filter: (a) creation of binary random vector (w = image width; $k = 20$); (b) convolution with a Gaussian kernel (an “SF Bubble”); (c) log-scaling; (d) construction of a 2D filter. (4) SF filtering by dot-multiplying the 2D filter with the stimulus’ FFT amplitude. (5) Inverse Fourier Transform. (6) Cropping.

of the technique in Experiment 1.

On each trial, the SF information of a stimulus was sampled randomly as illustrated in Figure 2. First, the square base stimulus was padded to minimize edge artifacts in the SF domain. It was centered on a uniform gray background of the stimulus' average luminance and of twice its size. In Experiment 1, for example, the plaid of size 256 x 256 pixels was padded with a mid-luminance background of size 512 x 512 pixels. Second, the padded stimulus was Fourier transformed using functions from the Image Processing Toolbox for Matlab (Mathworks). The quadrants of the Fourier image were shifted so that low SFs occupied the central region of the complex (i.e., real + imaginary number) amplitude matrix. Third, a random filter was constructed. The construction of this filter involved the following steps: (a) A binary random vector of $2wk$ elements was created, where w was the width of the base stimulus (multiplied by 2 for the padding) and k a constant that determined the smoothness of the sampling (the higher k , the smoother); k was arbitrarily set to 20 for all the experiments reported in this thesis.² In Experiment 1, the random vector thus had 10,240 elements (256 x 2 x 20). The vector contained zeros among b ones that were randomly distributed (with repetition). b determined the number of SF bubbles (see below) and was set to 45.³ (b) In order to create a smooth filter, the binary vector was convolved with a Gaussian kernel, referred to as an "SF bubble". The standard deviation of the SF bubble was arbitrarily set to 1.5 and the maximum to 0.125;

² The parameters reported in this thesis were determined through visual inspection of sampling vectors during the simulation of 20,000 SF Bubbles trials. The goals were to find settings that resulted in smooth filtering profiles that did not reveal too much or too little information relative to the bandwidth of SF channels in the visual system (e.g., see de Valois & de Valois, 1990, for a review). The parameter k can be used to increase or decrease the resolution of the filtering profile, thereby acting as a smoothing parameter.

³ The number of Bubbles was determined so that less than 1% of the simulated trials resulted in values larger than 1.0 in the sampling vector. This is important since multiplying by values greater than 1.0 would result in more energy at the respective SFs than present in the base stimulus. The number of Bubbles was kept constant in all experiments, and for the few trials on which values exceeded 1.0, we set them to 1.0.

all values of the resulting vector above 1.0 were set to 1.0.⁴ The convolution resulted in a “sampling vector” consisting of b randomly located SF bubbles. (c) To ensure that the sampling vector approximately fit the sensitivity of SF channels in the human visual system that were found to be tuned more narrowly to low than high SFs (see De Valois & De Valois, 1990, for a review), the smoothed vector was subjected to a logarithmic transformation: w elements of the vector were sampled according to the function of $e^{\frac{(x-1)\ln(kw)}{(w-1)}} + a$, with $x = [1:w]$ and $a = kw/2$. The constant a prevented low and high SFs being sampled less often than intermediate SFs. In Experiment 1, the image width w was 256 pixels and k equaled 20, thus resulting in $a = 2560$. (d) The w -element sampling vector was rotated about its origin to create a random 2-dimensional filter of size $w \times w$.

Fourth, filtering was carried out by dot-multiplying the 2-dimensional filter with the complex amplitude of the padded base stimulus, before subjecting the result to the inverse Fourier transform. We constructed the experimental stimuli by cropping the central $w \times w$ pixel region of the filtered image. Gaussian white noise was added to the SF sampled stimulus to adjust performance between floor and ceiling.⁵ The $w \times w$ noise field was multiplied by $1-c$ with c ranging from 0.0 to 1.0 and then added to the image multiplied by c . The initial value of c resulted from an informed guess (e.g., 0.76) and c was then increased or decreased on a block-by-block basis by the experimenter (in increments of .2; Experiment 2a, 2b, and 3) or on a trial-by-trial basis using QUEST to

⁴ The two criteria for choosing sigma were to minimize ringing artifacts in the output image that occur if the filter (i.e., the bubble) is too steep, and to not reveal too much information at once. From previous studies it is known that the SF bandwidths of individual channels in the visual system are mostly in the range of 1.0-1.5 octaves (De Valois et al., 1982; Wilson & Wilkinson, 1997).

⁵ Adding external noise, i.e. Gaussian white noise which has a zero correlation with the signal, is thought to not alter a task qualitatively since even under noiseless viewing conditions observers' performance is affected by internal noise; findings suggest that observers use the same strategy at all levels of external noise (e.g., Murray & Gold, 2004).

maintain a given performance level (Watson & Pelli, 1983) (Experiment 1 and 4). For example, in Experiment 1, performance was maintained at 75% correct.

To find out which SFs drove the observers' correct/incorrect responses or response times, a multiple linear regression was performed on the random binary vector (see above) and transformations of the observers' correct/incorrect responses or response times. Here, a multiple linear regression is linearly related to summing all sampling vectors weighted by the transformed responses. Throughout this thesis, correct/incorrect responses were transformed as follows: correct responses were given a value of $1 - P(\text{correct})$ —where $P(\text{correct})$ is the probability of a correct response—and incorrect responses a value of $-P(\text{correct})$. In Experiment 1, for example, correct responses were assigned a value of .25 and incorrect responses a value of -0.75. This transformation was done to make conditions with different accuracy rate comparable. Similarly, fast responses (RTs smaller than the median RT of the corresponding block) were given a value of 1 and slow responses (those greater than the median RT) were given a value of -1. For the RT analysis, only correct trials were used.

The vector of w regression coefficients—referred to as a “classification vector”—was Z -transformed for each observer. A group classification vector can be computed by summing the classification vectors of all observers and dividing the resulting vector by \sqrt{n} , with n equal to the number of observers. A pixel test was used to determine a statistical threshold (Chauvin et al., 2005). Note that due to padding, the x th element of the classification vector corresponds to $x/2$ cycles per base stimulus width; all SFs will always be given relative to the base stimuli in this thesis.

Results and discussion

The first block was considered as a practice block, and it was excluded from the analysis. Results are shown in Figure 3 which plots the Z -scores across all SFs (classification vectors) as well as the significance threshold ($p < .05$; $S_r = 256$; $FWHM = 3.53$; $Z_{crit} = 3.45$; for details, see Chauvin et al., 2005). Individual and group results are very similar (r 's = 0.93, 0.99, and 0.96); we will thus only discuss the group results. Two significant peaks occurred: the first peak at 10 cpi ($Z_{max} = 10.50$) and the second one at 45 cpi ($Z_{max} = 6.87$), with an octave width of 0.42 and 1.39, respectively. This demonstrates that the SF Bubbles method can reveal accurately the SFs that drive the observers' responses.

Experiment 2

Experiment 2 was designed to investigate which SFs underlie the identification of upright and inverted faces. In Experiment 2a, accuracy was used as measure of effective identification, whereas reaction time (RT) was used in Experiment 2b. Both versions of the experiment were based on the SF Bubbles method described above.

Methods

Six healthy University of Victoria students (four females, two males) aged between 19 and 35 years ($M = 24.6$ years) took part in the accuracy version or the RT version of the experiment. All observers had normal or corrected-to-normal vision. Five participants were naïve as to the purpose of the study whereas the sixth was the author. All participants completed the same practice phase, and four of them completed both Experiment 2a and 2b. The fifth participant in Experiment 2a was not available for the follow-up, so another subject participated only in Experiment 2b.

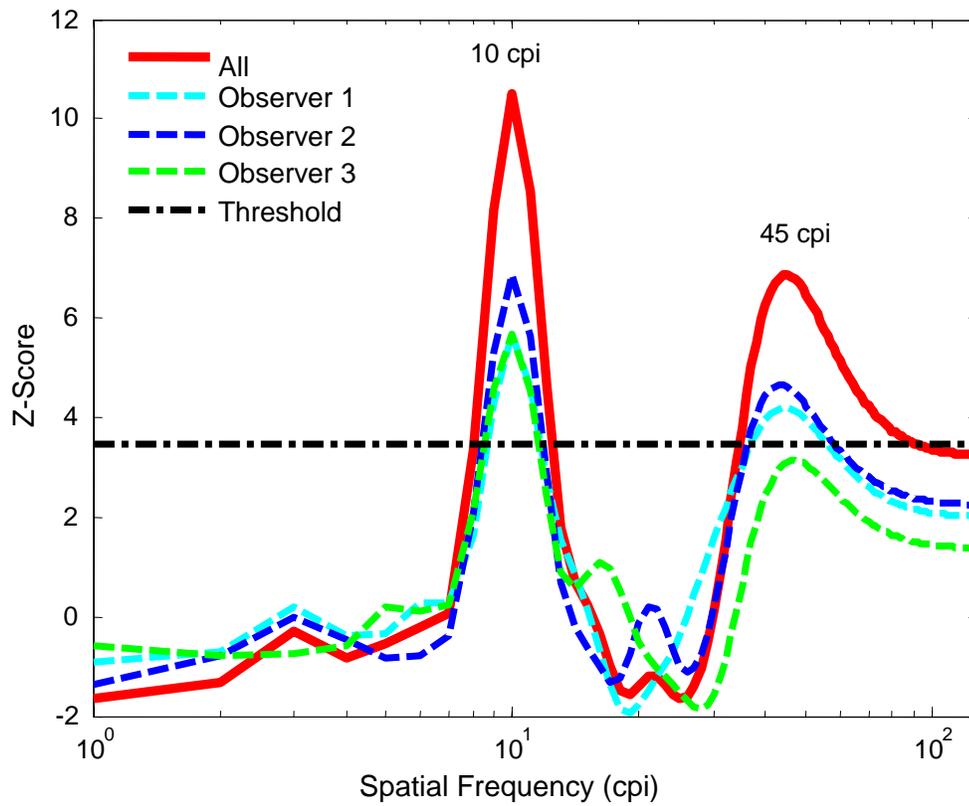


Figure 3. Individual and group classification vectors obtained in the plaid detection task (Experiment 1). The SF Bubbles technique revealed significant peaks at 10 cycles per image (cpi) and at 45 cpi ($Z_{crit} = 3.45$, $p < 0.05$), thus accurately showing the diagnostic SFs.

Twenty grayscale photos of faces of 256 x 256 pixels served as base stimuli. More specifically, five male and five female faces were shown, each with a neutral and a happy expression. We used two exemplars of each identity to make it less likely that observers followed a template-matching strategy. The main facial features (eyes, nose, mouth, chin, and forehead) were aligned as much as possible across the stimulus set. Faces were shown through an oval aperture in a mid-gray field so that only internal facial features were revealed (Figure 4). Mean luminance, contrast, and power spectra were equated as much as possible across face stimuli. On average, face width subtended a visual angle of 6.48° . The base stimuli were presented upright or inverted (rotated 180° in the image plane).

In the initial learning phase, participants were given printed pictures of the 20 faces (two exemplars of the same identity per page) with the corresponding name (e.g., Mary) on top. They were asked to learn to associate the faces with the names in both the upright and inverted orientation. The learning phase was self-paced and no further instructions as to how to learn the faces were given. When the participants reported that they could identify all faces correctly, the computerized practice sessions began. Each session lasted two hours per day including frequent rest breaks. First, participants completed the practice with upright and then with inverted faces. Participants were seated in a dark room and a chin rest maintained them at a 53 cm viewing distance. Each trial began with a central fixation cross presented for 435 ms, followed by an upright or an inverted face presented for 435 ms, and then by a homogeneous mid-gray field that remained on the screen until the observer responded by pressing the appropriate key on a computer keyboard. Each of the keys (numerals 0 to 9) was associated with a particular

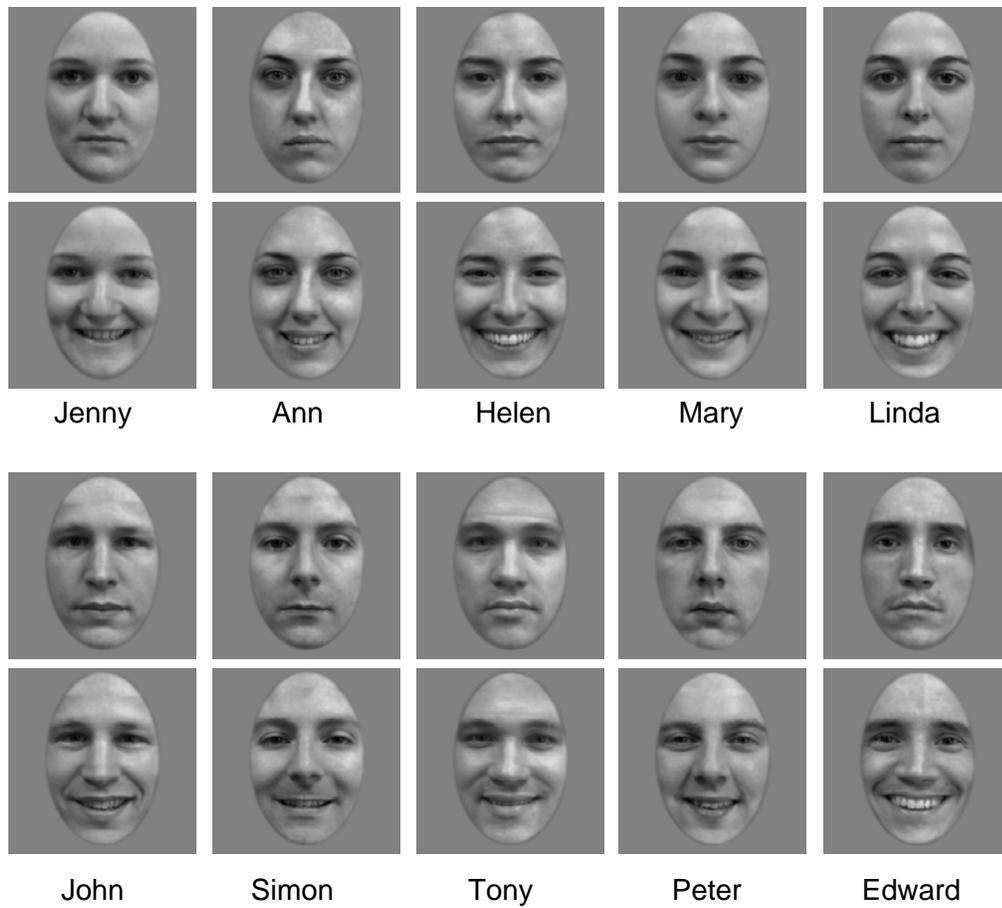


Figure 4. Base face stimuli used in Experiments 2a, 2b, and 3 with corresponding names that participants learned. Stimuli displayed ten identities (five males, five females) with two facial expressions (neutral, happy) each.

face name. When participants responded incorrectly, auditory feedback was provided (a brief 3000 Hz pure tone). The first part of the practice session was completed when accuracy for upright faces was above 95% for two successive blocks of 100 trials; the second part was completed when the same criterion was reached for inverted faces. On average, participants needed 6.4 practice blocks in the upright condition and 17.0 blocks in the inverted condition. Finally, participants performed three additional practice blocks with upright faces and three with inverted faces in which Gaussian white noise was added to the full-SF-spectrum faces. This was done in preparation for the experimental blocks (Experiment 2a and 2b) and to give the experimenter an idea about the initial amount of noise needed to reach the desired performance level.

SF Bubbles paradigm

In the experimental phase of the accuracy version (Experiment 2a), each participant was presented with a total of 2,100 upright and 2,100 inverted face stimuli. Upright and inverted faces were presented in separate 100-trial blocks, starting with an “upright” block and then alternating with “inverted” blocks. Testing sessions lasted two hours (one per day spread out over a week) with frequent breaks between 100-trial blocks. Accuracy was measured in the same 10-choice identification task as during the practice. The experimental trials differed from the practice trials as follows: (1) the SFs of the base stimuli were sampled (Figure 5; for details, see the *Spatial Frequency Bubbles* section of Experiment 1); (2) no feedback was given; and (3) performance in the upright blocks was maintained between 75% and 85% correct by adjusting the quantity of additive Gaussian white noise block per block. The same amount of noise was used in the following inverted blocks. We chose to equate the quantity of additive noise across

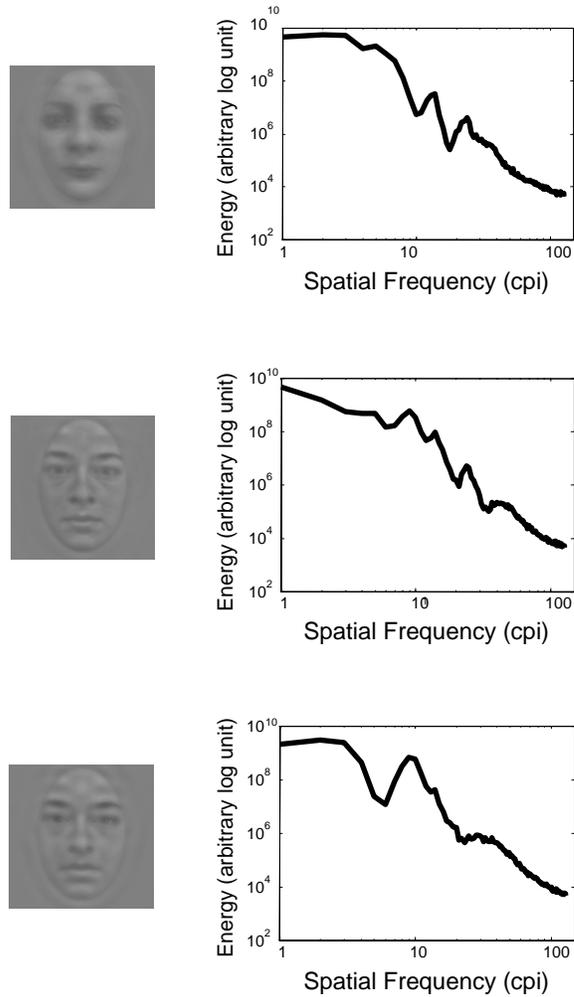


Figure 5. SF sampled output. Shown is a sample stimulus (Mary, see Figure 4) after sampling with SF Bubbles on three hypothetical trials. The graphs display the stimulus' Fourier amplitude averaged across orientations as a function of SF expressed in cycles per image (cpi).

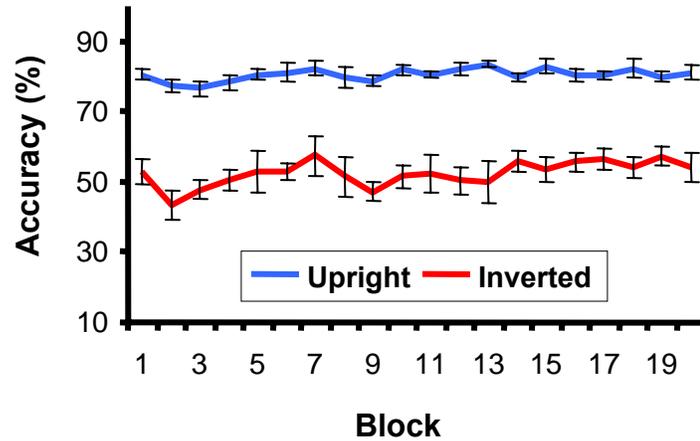
conditions instead of accuracy (the latter is what Sekuler et al., 2004, and Gaspar et al., 2005, did) because naturally upright and inverted faces contain the same amount of information. This way, all low-level information (except phase) was identical in the two conditions.

The RT version (Experiment 2b) followed Experiment 2a and differed from it only in four respects: (1) face stimuli remained on the screen until a response was made; (2) participants named aloud the identities of the faces, and a voice key was used to measure response latency; (3) after each trial, the experimenter typed the participant's response using a computer keyboard; and (4) performance was maintained above 90% correct in the upright condition block per block.

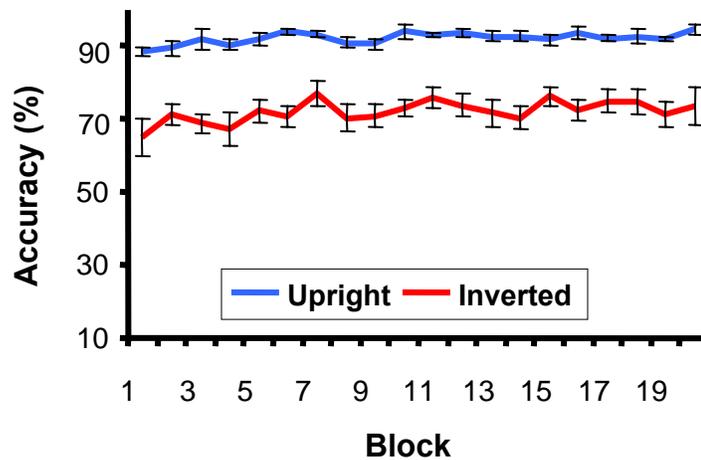
Results and discussion

In both the accuracy and the RT version, the first block in each orientation condition was excluded from the analyses. In Experiment 2a, accuracy across participants was significantly higher for upright ($M = 80.52\%$, $SD = 1.35$) than for inverted faces ($M = 52.43\%$, $SD = 6.32$), $t(4) = 8.56$, $p < 0.01$. In Experiment 2b, where accuracy was adjusted to above 90% for upright faces ($M = 92.24\%$, $SD = 2.23$), accuracy for inverted faces was again lower ($M = 72.12\%$, $SD = 6.12$), $t(4) = 6.81$, $p < 0.01$. Furthermore, RTs on correct trials were significantly shorter in the upright ($M = 1479.83$ ms, $SD = 315.29$) than in the inverted condition ($M = 1935.01$ ms, $SD = 410.95$), $t(4) = -6.90$, $p < 0.01$. Thus, both versions of the experiment exhibited a clear FIE. Figure 6 shows the accuracy and RTs over the 20 blocks of the experiments. The present finding of a stable FIE over the course of the experiment is consistent with results by Robbins & McKone (2003). They demonstrated that even after substantial training with inverted faces, holistic

(a) Exp. 2a



(b) Exp. 2b



(c) Exp. 2b

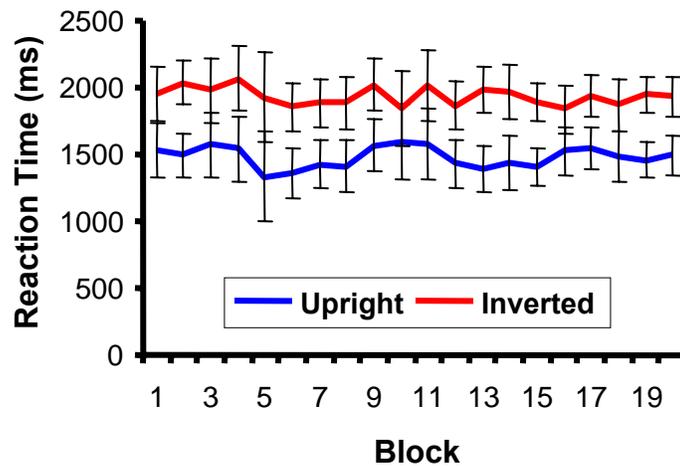


Figure 6. Mean accuracy over the 20 blocks of (a) Experiment 2a (“Accuracy Version”) and (b) Experiment 2b (“RT Version”). (c) Average reaction times of correct trials over the 20 blocks of Experiment 2b. Error bars give the standard errors of the mean. Both accuracy and RT results show a clear FIE.

processing (e.g., as indexed by the composite effect) was not learned for inverted faces, and concluded that orientation-specific face processing is highly stable against practice.

To reveal the SF ranges that led to accurate (Experiment 2a) and fast (Experiment 2b) face identification in the upright and inverted condition, we performed multiple linear regressions on the sampling vectors for each orientation condition per observer and for each regressor (see the *Spatial Frequency Bubbles* section of Experiment 1). The group classification vectors for the upright and inverted conditions and their normalized differences are shown in Figure 7 for Experiment 2a and Figure 8 for Experiment 2b, along with a sample face containing only the significant SF information revealed for the respective condition.

Since individual and group classification vectors were very similar in both experiments (average correlations including both orientation conditions of $r = 0.86$ in Experiment 2a and $r = 0.77$ in Experiment 2b), we will only report the group results. The upright group classification vector in Experiment 2a showed a significant SF band of 2.00 octaves and dual peaks, one at 7.14 cpf ($Z_{max} = 8.20$; $p < .05$; $S_r = 256$; $FWHM = 3.53$; $Z_{crit} = 3.45$) and the other at 12.14 cpf ($Z_{max} = 8.10$). Similarly, in the inverted condition, a 2.00 octaves wide SF band peaking at 7.14 cpf ($Z_{max} = 8.68$) and 11.07 cpf ($Z_{max} = 7.72$) was significant. In Experiment 2b, the group classification vector for the upright condition revealed a significant SF range of 1.78 octaves peaking at 8.57 cpf ($Z_{max} = 6.67$) and 12.86 cpf ($Z_{max} = 5.58$). In the inverted condition, a SF range of 0.73 octaves peaking at 12.14 cpf ($Z_{max} = 4.33$) was significant. None of the difference classification vectors reached statistical significance.

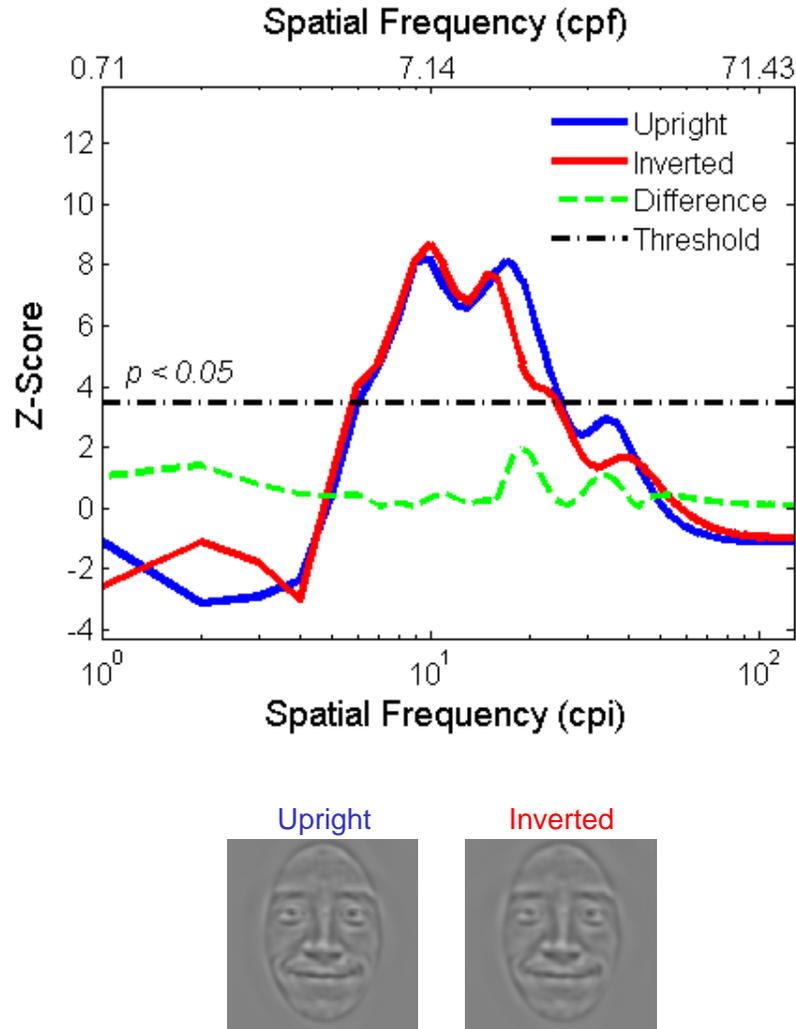


Figure 7. SF Bubbles results of the accuracy version of the face identification task (Experiment 2a). Z-scores are plotted as a function of SF in cycles per image (cpi; bottom x-axis) and cycles per face width (cpf; top x-axis). The group classification vectors for both upright and inverted faces exceeded the significance threshold of $Z_{crit} = 3.45$ ($p < 0.05$) for a SF band between 4.29 cpi and 17.14 cpi. No significant difference between the upright and inverted classification vectors was found ($p > 0.05$). The images at the bottom illustrate the significant SF information used in the upright (left) and inverted (right) conditions on a sample face.

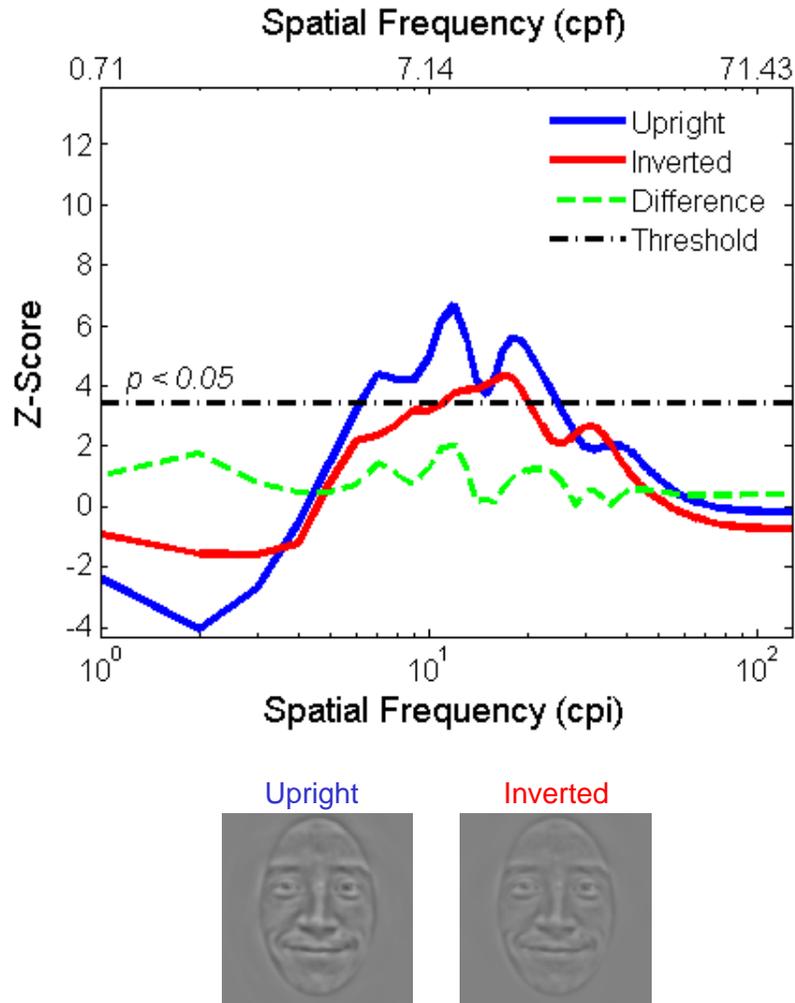


Figure 8. SF Bubbles results of the reaction time version of the face identification task (Experiment 2b). Z-scores are plotted in function of SF in cycles per image (cpi; bottom x-axis) and cycles per face width (cpf; top x-axis). The group classification vector in the upright condition exceeded the significance threshold of $Z_{crit} = 3.45$ ($p < 0.05$) for a SF band between 5.00 cpf and 17.14 cpf and in the inverted condition between 8.57 cpf and 14.29 cpf. No significant difference between the upright and inverted classification vectors was found ($p > 0.05$). The images at the bottom depict only the significant SF information used in the upright (left) and inverted (right) condition.

The correlations between the two versions of the experiment were very high in both the upright condition ($r = 0.96$) and the inverted condition ($r = 0.93$). Most importantly, the correlations between the classification vectors for upright and inverted faces was very high in both Experiment 2a ($r = 0.97$) and Experiment 2b ($r = 0.95$), confirming that the same SF band was used for identifying upright and inverted faces.

Overall, the results show no difference in the SF information used for upright and inverted face identification, but better performance in the upright condition. These findings are in accordance with the quantitative difference hypothesis.

Experiment 3

All variants of the Bubbles technique are designed to uncover the interaction between the available information (i.e., in the outside world) and the bottom-up (stimulus-driven) and top-down (goal-driven) visual strategies used by the observers (Gosselin & Schyns, 2002; Gosselin & Schyns, 2004; Murray & Gold, 2004). The results of Experiment 2 reveal no differences in SF tuning between the upright and inverted conditions. Based on Experiments 1 and 2 alone, we cannot rule out the possibility that this null result might be due to a relative insensitivity of the SF Bubbles approach to subtle differences in SF tuning resulting from bottom-up or top-down alterations in the visual strategies employed by observers (e.g., holistic vs. featural processing). In order to test this possibility, we carried out two follow-up experiments (Experiment 3 and 4) that addressed bottom-up and top-down induced differences in SF tuning, respectively.

Experiment 3 was based on the finding that the retinal projection of an image, and with it the availability of visual information, changes with viewing distance (e.g., Sowden & Schyns, 2006). For example, we might bring a newspaper closer to our eyes to read it,

whereas we might take a step back from a mosaic to get its global impression. When a face moves farther away from an observer, its retinal image size decreases, and progressively lower SFs are lost to the observer. Inversely, when getting closer, the retinal image size increases and the range of SFs available to resolve a task shifts to higher SFs in a bottom-up fashion (e.g., Loftus & Harley, 2005; Chung et al., 2002; Majaj et al., 2002). Experiment 3 examined whether we could reveal such a bottom-up induced SF tuning change with the SF Bubbles technique. Observers were asked to identify the same faces as in Experiment 2 but viewed at different visual angles (i.e., BIG and SMALL retinal image sizes). The hypothesis was to find a shift to lower SFs in the SMALL condition.

Method

One male and two female volunteers, aged between 22 and 26 years ($M = 23.7$ years), were recruited. All observers had normal or corrected-to-normal vision. The three participants were naïve as to the purpose of the study, and they received course credits or were paid for participating.

The base stimuli of Experiment 2 were used but their resolution was reduced to 128 x 128 pixels (in both the SMALL and BIG conditions). In the SMALL condition, observers saw the upright face stimuli at a screen resolution of 2048 x 1536 pixels and at a distance of 180 cm, resulting in face width of 0.45° of visual angle. In the BIG condition, they saw them at a screen resolution of 640 x 480 pixels and at a viewing distance of 45 cm, resulting in a face width of 5.85° of visual angle.⁶

⁶ The changes in screen and stimulus resolution were made to create two conditions that varied sufficiently in visual angle while they were of similar difficulty and used stimuli with the same number of pixels. The settings allowed us to conduct Experiment 4 in the same (rather small) testing room as Experiments 1-3.

Participants performed the same practice and face identification task as in Experiment 2a with upright faces. Each participant completed 3,300 trials per condition in two-hour sessions (one per day spread out over a week) with frequent breaks between blocks. We adjusted the quantity of additive Gaussian white noise as described in the methods section for Experiment 1 on a block-by-block basis so that performance was approximately 80% correct in the BIG condition (i.e., the easiest condition). The 100-trial blocks succeeded each other as in Experiment 2a, but this time alternating between BIG and SMALL rather than between upright and inverted.

Results and discussion

For each participant, the first block of each task was discarded from the analysis. The analysis was thus performed on 19,200 trials (3,200 trials per size condition x 2 size conditions x 3 participants). Since individual and group results were very similar (with an average correlation of $r = 0.93$ including both conditions), we will only report group results.

Mean accuracy in Experiment 3 was similar in the SMALL ($M = 74.48\%$, $SD = 9.79$) and BIG conditions ($M = 80.36\%$, $SD = 1.84$), $t(2) = 1.227$, $p > 0.05$.

In the SMALL condition, a range of SFs of octave width 1.8 peaking at 5 cpf ($Z_{max} = 12.46$) exceeded the significance threshold ($p < .05$; $S_r = 128$; $FWHM = 3.53$; $Z_{crit} = 3.25$). Note that the significance threshold (Z_{crit}) is slightly lower than in Experiment 2a because of the reduced stimulus resolution ($S_r = 128$ pixels instead of $S_r = 256$ pixels). In the BIG condition, an SF range of octave width 2.81 peaking at 8.57 cpf ($Z_{max} = 9.19$) attained significance. The difference between the group classification vectors of the two conditions reached significance between 3.57 cpf and 5.00 cpf as well as between 9.29

cpf and 17.86 cpf with a maximum at 4.29 cpf ($Z_{max} = 5.27$). Thus, the hypothesized shift towards lower SFs with smaller retinal image size occurred, as can be seen in Figure 9.

This shift to coarser information in the SMALL condition is in accordance with the findings by Majaj et al. (2002) and Chung et al. (2002) for letter identification. It is also in accordance with the results of Näsänen (1999) who employed two different viewing distances in one of his face identification tasks. Importantly, the findings of Experiment 3 show that the SF bubbles approach is capable of revealing changes in SF tuning based on the same task and the same stimuli as in Experiment 2. It would be interesting to investigate further how different tasks are performed over different retinal image sizes to test the idea that some recognition tasks can be performed over a range of distances while others cannot (e.g., Sowden & Schyns, 2006).

Experiment 4

Experiment 3 showed that SF Bubbles can reveal differences due to changes in the availability of visual information (bottom-up). Since in Experiment 2 the availability of visual information was equated between the upright and inverted conditions, it is also important to assess the capacity of the SF Bubbles technique to reveal subtle top-down induced differences in SF tuning to see whether the null result is real. In Experiment 4, we modified task demand—a top-down factor—while stimulus information remained identical in all conditions. The SF Bubbles technique was applied to an expression (happy vs. neutral) and a gender discrimination task that have previously been shown to induce different SF usage patterns (e.g., Gosselin & Schyns, 2001; Schyns & Oliva, 1999). Using hybrid stimuli (i.e., a low-pass filtered face superimposed on a high-pass filtered face) Schyns and Oliva revealed a bias towards high SFs for an expressive vs. non-

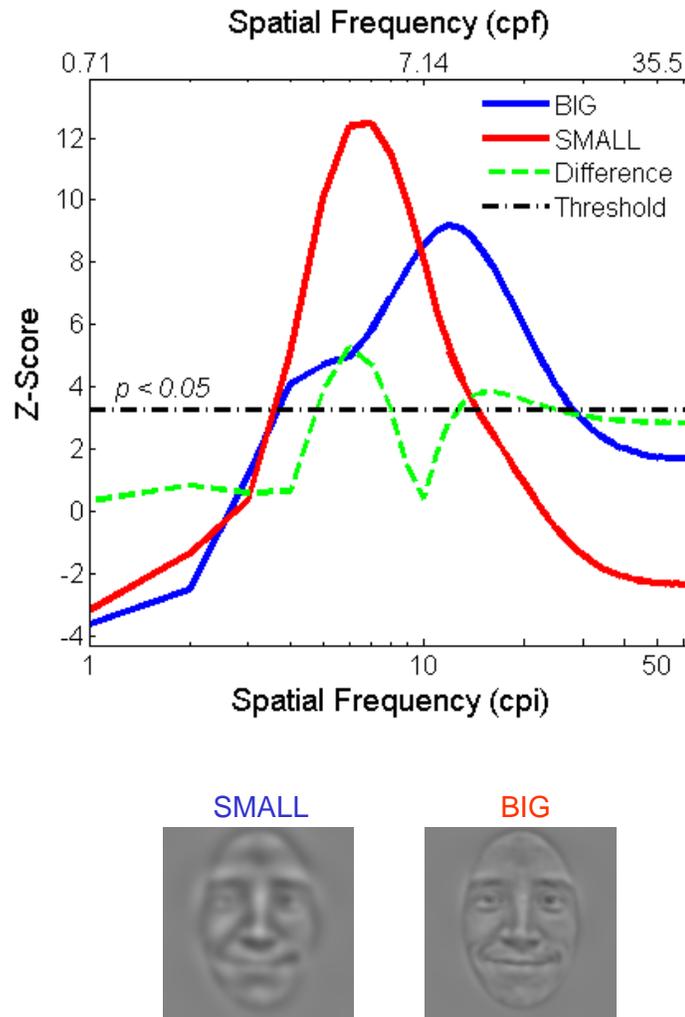


Figure 9. SF Bubbles results of Experiment 3. Z-scores are plotted in function of SF in cycles per image (cpi; bottom x-axis) and cycles per face width (cpf; top x-axis). In the BIG condition, the group classification vector peaked significantly at 8.57 cpf and in the SMALL condition at 5 cpf ($Z_{crit} = 3.25$; $p < 0.05$). The difference between the SMALL and BIG classification vectors reached significance—the SF Bubbles technique revealed the hypothesized shift to lower SFs with smaller stimulus size. The bottom pictures show only the significant SFs used in the SMALL (left) and BIG (right) condition.

expressive categorization task, a low-SF bias for an expression identification task (happy, neutral, or angry) and neither a high- nor a low-SF bias for a gender categorization task. Our main purpose was not to investigate the specific SFs underlying the different tasks but to provide an “existence proof” that SF Bubbles can reveal a flexible SF use with varying top-down factors while bottom-up factors remained identical.

Method

We recruited 40 healthy undergraduate students (31 females, 9 males) aged between 18 and 42 years ($M = 19.8$ years). They all had normal or corrected-to-normal vision, were naïve as to the purpose of the experiment, and received course credit for participating.

For the gender and happy vs. neutral discriminations, the same ten neutral faces (five males, five females) were used as in Experiment 2. The corresponding ten happy faces, however, differed from the set previously used in that no teeth were visible (Figure 10). We chose this set to make task difficulty between gender and happy vs. neutral discriminations more similar; a pilot study showed ceiling effects for happy vs. neutral discrimination when accuracy for gender was in the 65-75% range. Base stimuli had a resolution of 256 x 256 pixels and face width subtended a visual angle of 6.48°. They were normalized for a number of low-level visual features and for the main facial feature position as in Experiment 2. In preparation for the experiment, participants viewed all original faces briefly prior to the experiment on a handout.

Experiment 4 was divided in two parts: each participant completed six consecutive 100-trial blocks of happy vs. neutral discrimination and six consecutive 100-trial blocks of gender discrimination. The first 20 participants initially completed the



Figure 10. Base face stimuli used in Experiment 4 for the happy vs. neutral and gender discrimination tasks. The stimulus set consisted of ten male and ten female faces, each with a neutral and a happy expression. The neutral faces were the same as those used in previous experiments but the happy faces differed slightly from those previously used in that teeth were not visible.

happy vs. neutral discrimination, followed by the gender discrimination; the last 20 participants completed the tasks in opposite order. The experiment was completed by all participants within a one-hour session with breaks between blocks. Each trial began with a central fixation cross presented for 412 ms, followed by an upright face presented for 412 ms, and then by a homogeneous mid-gray field that remained on the screen until the observer responded by pressing the appropriate key on a computer keyboard. Stimulus duration was chosen to be slightly shorter than in Experiment 2 because of the simplified nature of the task (categorization with 2-alternatives vs. individuation with 10-alternatives). Keys were counter-balanced across participants. For the initial task, performance was adjusted on a trial-by-trials basis by manipulating the quantity of additive noise using QUEST (Watson & Pelli, 1983). In the second task in Experiment 4, the same experimental stimuli were used (i.e., same base stimuli, same sampling vector, and same Gaussian white additive noise) as in the corresponding trials of the first task in Experiment 4. This ensured that stimulus information was exactly the same in both tasks.

Results and discussion

For each observer, the first block per task was excluded from the analysis. Thus, the following results are based on 40,000 trials in total (500 trials per task x 2 tasks x 40 participants). Since Experiment 4 was based on a large number of observers who each completed a relatively small number of trials, we will only report group results. Mean accuracy across participants was significantly higher for happy vs. neutral ($M = 81.20\%$, $SD = 5.20$) than for gender discriminations ($M = 65.31\%$, $SD = 5.03$), $t(39) = 16.899$, $p < 0.001$.

The group classification vectors for happy vs. neutral and gender discriminations are illustrated in Figure 11. For each task, two separate peaks reached significance ($p < .05$; $S_r = 256$; $FWHM = 3.53$; $Z_{crit} = 3.45$). For the gender discrimination, the peaks occurred at 2.86 cpf ($Z_{max} = 7.11$) and 7.14 cpf ($Z_{max} = 8.16$). The classification vector for happy vs. neutral discrimination peaked at 2.86 cpf ($Z_{max} = 5.87$) and at 6.43 cpf ($Z_{max} = 5.70$). The correlation between the classification vectors for the two conditions was high ($r = 0.92$) and the difference, largest at 8.57 cpf ($Z_{max} = 2.90$), only approached significance.

This pattern of results was surprising in two respects: first, two separate SF ranges were revealed for each of the tasks which could not easily be predicted from the results obtained with hybrid stimuli by Schyns and Oliva (1999), and second, we did not find a difference in SF use for happy vs. neutral and gender discrimination when looking at the results across all stimulus types. The present findings suggest that the same (two) SF channels were involved in both discriminations. Schyns and Oliva pointed out that SF use for category judgments is determined by a complex interaction of a number of factors (e.g., the exact nature of the stimuli, the task, and set effects), which—together with the different methods used—might explain why we did not replicate the differential low- and high-SF biases reported in their study.

The purpose of Experiment 4 was to reveal flexibility of SF use due to task demand alone. Even though the overall classification vectors did not reveal differences, it is possible that for a particular stimulus type, e.g., neutral female faces, the information diagnostic for the two tasks resides at different SFs. In order to investigate the nature of the two-peak pattern and to isolate the effect of task demand, we analyzed the results for

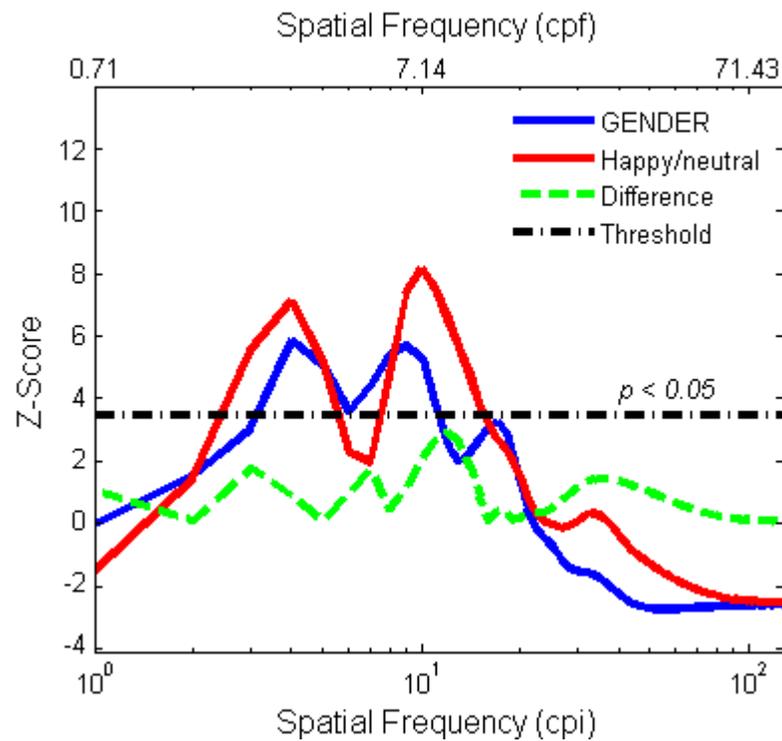


Figure 11. SF Bubbles results for the happy vs. neutral and gender discrimination tasks of Experiment 4. Z-scores are plotted in function of SF in cycles per image (cpi; bottom x-axis) and cycles per face width (cpf; top x-axis). For the happy vs. neutral discrimination, two significant peaks occurred (at 2.86 cpf and 7.14 cpf, $Z_{crit} = 3.45$; $p < 0.05$). Similarly, for the gender task, two peaks reached significance (at 2.86 cpf and 6.43 cpf). The difference between the classification vectors only approached significance.

the four stimulus types (happy female, happy male, neutral female, and neutral male) separately.

In this separate analysis, differences between the group classification vectors for the happy vs. neutral and gender discriminations were significant for all four stimulus types (Figure 12). With happy female faces, the normalized absolute difference between the classification vectors exceeded the threshold at 1.43 cpf ($Z_{max} = 3.91$; $p < .05$; $S_r = 256$; $FWHM = 3.53$; $Z_{crit} = 3.45$; for details, see Chauvin et al., 2005). The classification vectors for the two tasks peaked at the same SF of 2.86 cpf ($Z_{max} = 9.27$, happy vs. neutral; $Z_{max} = 11.75$, gender) and had a relatively high correlation of $r = 0.90$. With neutral female faces, the difference between the classification vectors for the two tasks was significant between 1.43 cpf and 5.71 cpf ($Z_{max} = 8.24$) as well as between 7.14 cpf and 10 cpf ($Z_{max} = 6.83$). Here, the correlation between group classification vectors for the two tasks was relatively small ($r = 0.39$) and two distinct peaks were revealed for happy vs. neutral (7.86 cpf, $Z_{max} = 9.33$) and gender discriminations (2.14 cpf, $Z_{max} = 7.84$). With happy male faces, the difference was significant between 1.43 cpf and 5.00 cpf ($Z_{max} = 9.14$). The reverse pattern to neutral female stimuli could be observed, in that the significant portion of the classification vector for happy vs. neutral was shifted into a lower SF range (with peaks at 2.14 cpf, $Z_{max} = 6.45$, and 5 cpf, $Z_{max} = 6.37$) than the peak for the gender task (7.86 cpf, $Z_{max} = 3.94$) with $r = -0.52$. With neutral male faces, the difference was significant between 1.43 cpf and 2.14 cpf ($Z_{max} = 3.84$). Happy vs. neutral peaked at 8.57 cpf ($Z_{max} = 7.75$), and gender at 10 cpf ($Z_{max} = 4.10$) with $r = 0.44$. The significant SF information for each stimulus type and task is shown in Figure 13.

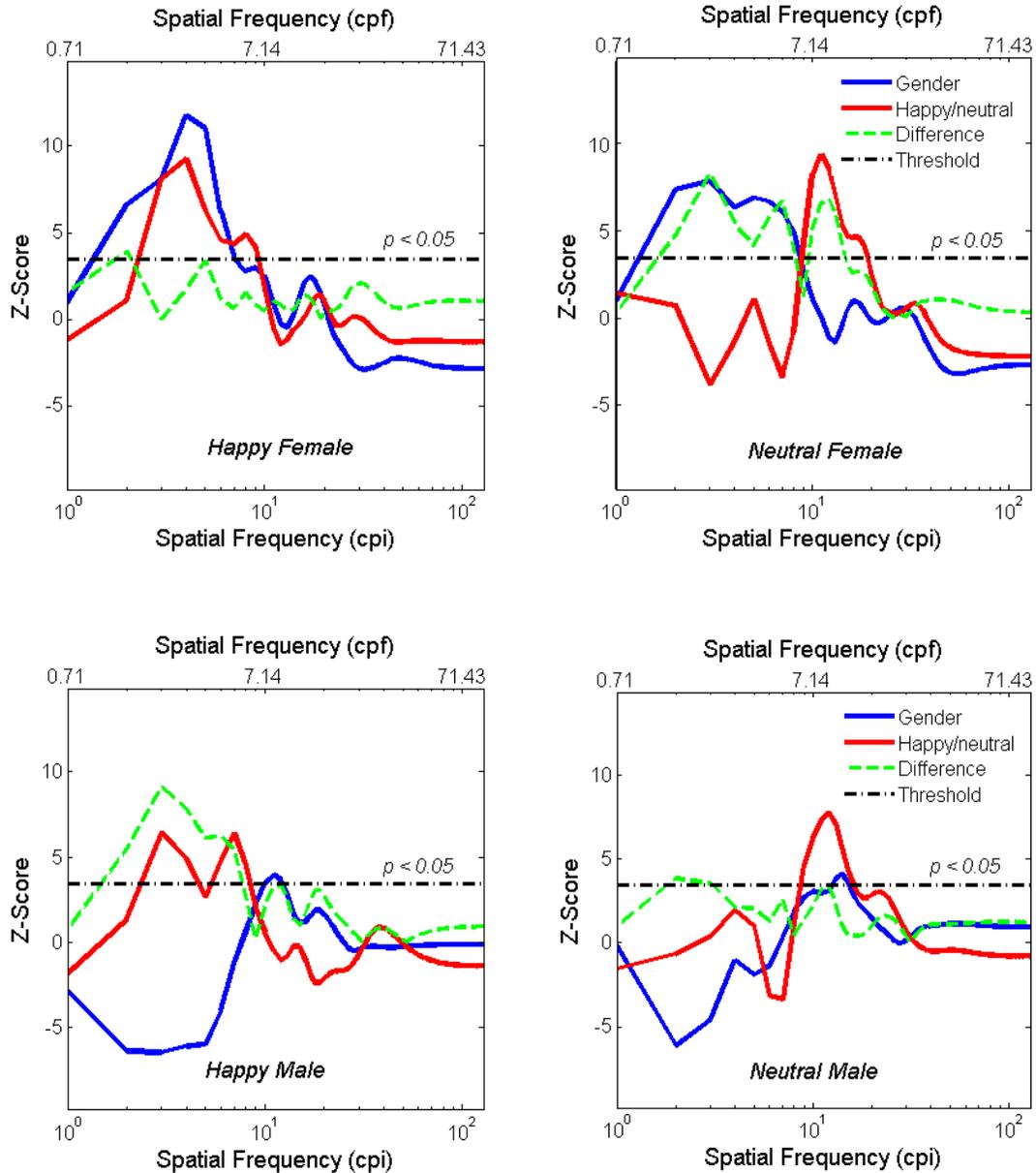


Figure 12. SF Bubbles results of Experiment 4 compared for the individual stimulus types: happy female (top left); neutral female (top right); happy male (bottom left); neutral male (bottom right). For all stimulus types, the SF difference between happy vs. neutral and gender discriminations was significant. The different peaks revealed for the two tasks, even though stimulus information was exactly the same, indicate that the SF Bubbles technique is sensitive to subtle top-down differences in SF tuning.

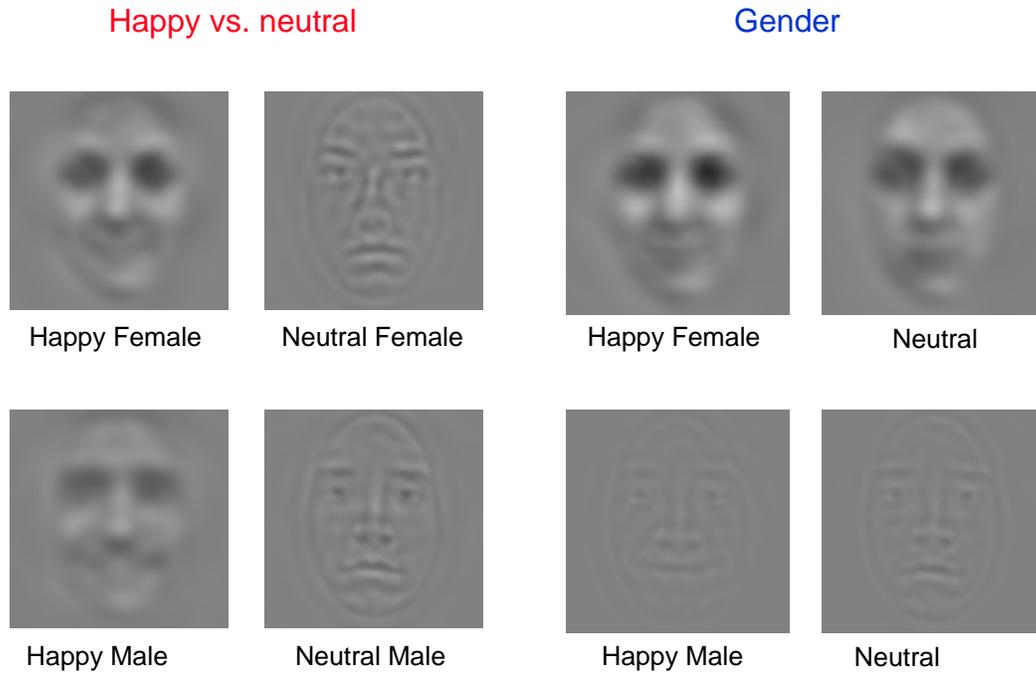


Figure 13. Illustration of the significant SF information used for the happy vs. neutral discrimination (left) and the gender discrimination (right) for all four stimulus types.

The results support the view by Schyns & Oliva (1999) that the information required for different tasks can be present at different SFs of the same stimulus, and that our visual system is flexibly tuned to extract this information. Furthermore, for both male and female faces, the happy vs. neutral classification vectors were only weakly correlated ($r = -0.03$ and $r = 0.22$, respectively), and for both happy and neutral faces, the gender classification vectors had a relatively weak correlation ($r = -0.46$ and $r = -0.48$). This finding is in accordance with results obtained by Smith, Cottrell, Gosselin, and Schyns (2005) who investigated which information human observers use for recognizing the six basic facial expressions (fear, happiness, sadness, disgust, anger, and surprise) in the spatial domain. Their results suggest that faces “evolved to send expression signals that have low correlations with one another and that the brain, as a decoder, further decorrelates and therefore improves these signals” (p. 188). Our results are in accordance with the view that the encoding and decoding of facial information is very efficient with minimal information overlap. Most importantly for the purpose of the present study, the results of Experiment 4 for the individual stimulus types show that SF Bubbles is sensitive to differences in SF tuning that are only due to a change in task demand.

Together the results of Experiment 3 and Experiment 4 show that the SF Bubbles approach is capable of revealing subtle differences in SF tuning for complex stimuli induced by retinal stimulus size—a bottom-up factor— (Experiment 3) and by task demands—a top-down factor—(Experiment 4). These results demonstrate the validity of the SF Bubbles technique (along with those of Experiment 1) and they suggest that the null result of Experiment 2 is real; thus, the FIE cannot be attributed to qualitative processing differences at the SF level.

General Discussion

The goal of the present study was twofold: (1) to introduce and validate a variant of the Bubbles technique (Gosselin & Schyns, 2001) capable of revealing the SFs that underlie the performance of human observers in a particular task and with a particular type of stimulus and (2) to use this technique to uncover the SFs upon which upright and inverted face identification are based. Compared to techniques previously used for a similar purpose—such as high-pass, low-pass, and band-pass filtering (e.g., Goffaux & Rossion., 2006), or critical band masking (e.g., Gaspar et al., 2005)—SF Bubbles minimizes the risk that observers adapt to a certain SF range during the task by randomly sampling the SF information on a trial-by-trial basis. Furthermore, the method makes no *a priori* assumption about the number of underlying SF channels but is capable of exploring all possible combinations of SFs in an unbiased way. The SF Bubbles technique is distinct from the application of the Bubbles technique in the spatial domain which operates by placing Gaussian windows on a given image that may be full-spectrum or SF filtered (e.g., Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002). SF Bubbles operates on the whole image and only samples the images' SF content trial per trial. Thus, it does not restrict the processing of a face to its local cues.

Three experiments were reported to validate the SF Bubbles technique. The first experiment demonstrated that the SF Bubbles method is capable of precisely revealing both low and high SFs that actually make up the stimulus in a detection task. Experiment 3 showed that the technique can reveal changes in SF tuning induced by subtle bottom-up influences. In particular, SF Bubbles was shown to be sensitive to differences in SF tuning resulting solely from changes in retinal image size. In accordance with previous

findings, the technique revealed a shift to lower SFs with smaller size (Chung, Legge, & Tjan, 2002; Loftus & Harley, 2005; Majaj et al., 2002; Näsänen, 1999). Experiment 4 demonstrated that SF Bubbles can reveal changes in SF tuning solely due to task demands—i.e., a top-down factor. Observers were presented with identical stimuli on corresponding trials of a gender discrimination task and a neutral vs. happy discrimination task. When isolating task demand by analyzing results for the four stimulus types separately (i.e., happy female, neutral male, happy male, and neutral female faces), significant differences were found in SF usage for the gender and happy vs. neutral discriminations. Flexible SF use was also found by Schyns and Oliva (1999) for different face categorization tasks (i.e., gender, expressive or not, which expression, and identity) based on hybrid images. The goal of Experiment 4 was not to investigate the precise SFs underlying the different categorization tasks, but to show that SF Bubbles is capable of revealing differences in SF use that can only be attributed to task demands. Together the results of Experiment 3 and 4 demonstrate that SF Bubbles is sensitive to both differences in SF tuning induced by top-down factors and bottom-up factors, thus suggesting that an obtained null result is real.

In the main experiment reported in this thesis (Experiment 2), the SF tuning in upright and inverted face identification was measured (with 10 identities x 2 exemplars) using SF Bubbles. While accuracy was 28% higher and reaction times 455 ms shorter with upright faces, thus showing a clear inversion effect, SF tunings were remarkably similar in both orientation conditions. In particular, a single SF band of ~2 octaves centered at ~9 cycles per face width was used. These results were observed independently on the accuracy (Experiment 2a) and the reaction time (Experiment 2b) classification

vectors. The findings of this study show that there are no qualitative differences at the SF level and are in accordance with a quantitative account of the face inversion effect (Sekuler et al., 2004; see also Gaspar et al., 2005). Even though efficiency was not measured directly in the present study, i.e. by comparing performance of human observers to an ideal observer, RT and accuracy results are in accordance with the view that the same SF information is processed more efficiently when faces are upright than inverted.

The present results are congruent with those from several previous experiments that assessed the SFs underlying face identification. Most of these experiments focused on the SFs used for upright faces and discovered a single SF channel 1.4 to 2 octaves wide with a peak between 8 and 16 cycles per face width, based on face stimuli subtending a horizontal visual angle between approximately 2.4° and 9.5° (e.g., Costen, Parker, & Craw, 1994, 1996; Gold, Sekuler, & Bennett, 1999b; Näsänen, 1999; Gaspar et al., 2005; for a review see Ruiz-Soler & Beltran, 2006). Here, the SF tuning found with upright faces was of ~ 9 cpf with an octave width of ~ 2 for stimuli that subtended a visual angle of $\sim 6^\circ$.

With inverted faces, however, few and mixed findings have been reported. Using critical band masking, Gaspar et al. (2005) revealed the same SF channel (~ 8 cpf) for the identification of upright and inverted faces. However, Nakayama's (2003) findings based on a similar approach but different task and stimuli suggested a broader SF channel for inverted than for upright faces. Thus more research was needed to elucidate SF use in upright and inverted face identification under different conditions. The results reported here replicate the findings of Gaspar et al. (2005) based on accuracy. Furthermore, our

results strengthen the conclusion of invariant SF use with upright and inverted faces by including RT analyses, by using an unbiased SF probing technique, and by using the same quantity of additive noise for upright and inverted faces. Here lies an important difference between our and Gaspar et al.'s (2005) approach—whereas they maintained performance at the same accuracy level by adjusting the amount of external noise independently in the upright and inverted conditions, we equated the quantity of noise in the two conditions and measured the FIE in terms of accuracy and RT. We chose this approach since naturally upright and inverted faces contain the same amount of information. The present findings show that controlling for the quantity of visual information does not change the outcome that the same SFs are used for upright and inverted faces.

The comparison between the present results and those of Goffaux and Rossion (2006) is not as straightforward. At first glance, the two sets of results appear contradictory. It is widely believed that upright faces are processed holistically and that inverted faces are processed featurally (e.g., Rossion, *in press*), and the results of Goffaux and Rossion suggest that different SFs carry the information relevant to holistic and featural processing (see also Sergent, 1986). In contrast, we reveal the same SFs for the processing of upright and inverted faces. However, the studies tested different hypotheses, and there are vast methodological differences between them. For example, Goffaux and Rossion (2006) examined whether there is a mapping between holistic/featural processing and low/ high SFs. For that purpose, they investigated the interaction between both the face composite effect and the part-whole advantage with low-pass and high-pass filtered faces maximized in difference (by choosing filtering

cutoffs that were two octaves apart). In the current study, we examined which SFs are optimal for upright and inverted face processing by using random filters that spanned the whole SF spectrum (from 1 cpi to the Nyquist frequency). Our results are difficult to compare with the findings of Goffaux and Rossion, because their study provides little direct insights into the interactions between the mid-SF range revealed in the present study and holistic/featural processing. In turn, the present study makes little direct predictions about the strategies or mechanisms with which the same information is processed beyond the SF encoding stage. Our results do not refute the qualitative difference hypothesis once and for all but they place constraints on where qualitative differences might be found.

Furthermore, findings by Cheung, Richler, Palmeri, and Gauthier (in press) indicate that the results reported by Goffaux and Rossion (2006) can be attributed to a response bias. Cheung et al. used essentially the same sequential matching task as Goffaux and Rossion: participants were asked to judge whether the top halves of aligned or misaligned composites (stimuli composed of the top half of one face and the bottom part of another face) were identical. They replicated Goffaux and Rossion's result of a greater alignment effect for low SF than high SF faces on trials where the to-be matched top halves were identical and the bottom parts were different. However, they also found a greater bias for responding "same" for high SF than low SF faces in general. When this response bias was taken into account, low SF and high SF faces were found to be processed equally holistically. Even though their study is not directly comparable to ours, these results help to reconcile the present findings with the ones obtained by Goffaux and Rossion (2006).

Together, the present results and the findings by Sekuler et al. (2004) show that we use SF information of ~ 9 cpf in the eye and eyebrow regions in a linear fashion for processing both upright and inverted faces, albeit not with equal efficiency. It remains an avenue for future research to investigate whether different or the same (expertise-related) mechanisms mediate the faster and more accurate processing of the diagnostic information in the upright than in the inverted condition.

Conclusion

Using a novel SF sampling technique, SF Bubbles, we showed that effective identification of upright and inverted faces relies on an SF band of ~ 2 octaves, centered at ~ 9 cycles per face width. Both accuracy and RT results indicate that the same SF information is processed more efficiently in the upright than in the inverted condition. The results thus suggest that the FIE cannot be attributed to early qualitative processing differences at the SF encoding stage.

References

- Biederman, I. & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London*, 352, 1203-1219.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433-436.
- Carey, S., Diamond, R., & Woods, B. (1980). The development of face recognition: A maturational component. *Developmental Psychology*, 16, 257-269.
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, 5, 659-667.
- Cheung, O. S., Richler, J. J., Palmeri, T. J., Gauthier, I. (in press). Revisiting the Role of Spatial Frequencies in the Holistic Processing of Faces. *Journal of Experimental Psychology: Human Perception and Performance*.
- Chung S. T. L., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research*, 42, 2137-2152.
- Collishaw, S. M. & Hole, J. C. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, 29, 893-909.
- Costen, N.P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception*, 23, 129-146.
- 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.
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545-559.

- De Valois, R. L. & De Valois, K. K. (1990). *Spatial Vision* 2nd edition (New York: Oxford University Press).
- Diamond, R. & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107-117.
- Fiset, D., Blais, C., Gosselin, F., & Schyns, P. (2006). Effective frequency tuning of three face categorization tasks [Abstract]. *Journal of Vision*, *6*(6):273, 273a, <http://journalofvision.org/6/6/273/>, doi:10.1167/6.6.273.
- 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*, 159-170.
- Gaspar, C. M., Sekuler, A. B., & Bennett, P. J. (2005). Upright & inverted face recognition relies on the same, narrow band of spatial frequencies [Abstract]. *Journal of Vision*, *5*(8):821, 821a, <http://journalofvision.org/5/8/821/>, doi:10.1167/5.8.821.
- Goffaux, V., Hault, B., Michel, C., Vuong, Q.C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, *34*, 77-86.
- Goffaux, V. & Rossion, B. (2006). Faces are “spatial”- holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1023-1039.
- Gold J., Sekuler A. B., & Bennett, P. J. (1999b). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, *39*, 3537-3560.
- Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, *28*, 167-207.

- Gosselin, F. & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition. *Vision Research*, *41*, 2261-2271.
- Gosselin, F. & Schyns, P. G. (2002). RAP: a new framework for visual categorization. *Trends in Cognitive Science*, *6*, 70-77.
- Gosselin, F. & Schyns, P. G. (2004). No troubles with Bubbles: A reply to Murray and Gold. *Vision Research*, *44*, 471-477.
- Gosselin, F. & Schyns, P. G. (2005). Bubbles: A user's guide. In L. Gershkoff-Stowe & D. H. Rakison (Eds.) *Building Object Categories in Developmental Time* (pp. 91-106). Hillsdale, NJ: Erlbaum.
- Hole, G. (1994) Configurational factors in the perception of unfamiliar faces. *Perception* *23*, 65–74.
- Leder H. & Bruce V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology*, *51*, 449–473.
- 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*, 513–536.
- Leder H., Candrian G., Huber O., Bruce V. (2001). Configural features in the context of upright and inverted faces. *Perception*, *30*, 73–83.
- Loftus, G.R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review*, *12*, 43-65.
- Majaj, N. J., Pelli, D. G., Kurshan P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, *42*, 1165-1184.
- Maurer, D., Le Grand, R. & Mondloch, C. J. (2002). The many faces of configural

- processing. *Trends in Cognitive Sciences*, 6, 255-260.
- McCotter, M., Gosselin, F., Sowden, P., & Schyns, P. G. (2005). The use of visual information in natural scenes categorization. *Visual Cognition*, 12, 938-953.
- Morrison, D. J. & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, 8, 454-469.
- Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2005). Classification images predict absolute efficiency. *Journal of Vision*, 5(2):5, 139-149.
- Murray, R. F., & Gold, J. M. (2004). Troubles with Bubbles. *Vision Research*, 44, 461-470.
- Murray, R. F., & Gold, J. M. (2004). Reply to Gosselin and Schyns. *Vision Research*, 44, 479-482.
- Nakayama, K. (2003). Face specific processing: role of local features in an affine metric [Abstract]. *Journal of Vision*, 3(9):91, 91a, <http://journalofvision.org/3/9/91/>, doi:10.1167/3.9.91.
- Näsänen R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, 39, 3824-33.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437-442.
- Reed, C. L., Stone, V. E., Bozova, S., Tanaka, J. (2003). The body-inversion effect. *Psychological Science*, 14(4) , 302-308.
- Robbins, R. & McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition*, 88(1), 79-107.
- Rossion, B., & Gauthier, I. (2002). How Does the Brain Process Upright and Inverted

Faces? *Behavioral and Cognitive Neuroscience Reviews*, 1, 62-74.

Rossion, B. (in press). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*.

Rossion, B. & Boremanse, A. (in press). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: evidence from the face composite illusion. *Journal of Vision*.

Ruiz-Soler, M. & Beltran, F. S. (2006). Face perception: An integrative review of the role of spatial frequencies. *Psychological Research*, 70, 273-292.

Scapinello, K. F. & Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science*, 21, 329-331.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, 13, 402-209.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69, 243-265.

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, 391-396.

Sergent, J. (1986). Microgenesis of face perception. In H. D. Ellis, M.A. Jeeves, F. Newcombe, & A. M. Young (Eds.), *Aspects of face processing* (pp. 17-33). Dordrecht: Martinus Nijhoff.

- Smith, M. L., Cottrell, G. W., Gosselin, F., Schyns, P. G. (2005). Transmitting and Decoding Facial Expressions. *Psychological Science, 16*, 184-189.
- Solomon, J. A. & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369*, 395–397.
- Sowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Sciences, 10*, 538-545.
- Tanaka, J. W. & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46A*, 225-245.
- Tanaka, J. W. & Sengco, J. (1997). Features and their configuration in face recognition. *Memory and Cognition, 25*, 583–592.
- Tanaka, J. W., Kaiser, M. J., Bub, D. N., & Pierce, L. (in preparation). Featural and configural information in face perception: Comparing apples and oranges.
- Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471–491.
- Valentine, T. & Bruce, V. (1986a). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica, 61*, 259-273.
- Watson, A. B. & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*, 113–120.
- Williams, C. C. & Henderson, J. M. (in press). The face inversion effect is not a consequence of aberrant eye movements. *Memory and Cognition*.
- Wilson, H. R. & Wilkinson, F. (1997). Evolving concepts of spatial channels in vision: from independence to nonlinear interactions. *Perception, 26*, 939-960.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental Psychology,*

81, 41-145.

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