COPY #1

Effects of Spatial Frequency Overlap on Face and Object Recognition

Charles Alain Collin

Department of Psychology

McGill University, Montreal

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

March, 2000

0-612-69989-7

ACKNOWLEDGMENTS

I wish to thank my thesis supervisor Avi Chaudhuri and my research partner Chang Hong Liu for their invaluable assistance at every stage in this work. I also wish to thank Stephane J.M. Rainville for his assistance in programming and Nikolaus F. Troje for his assistance in providing stimuli. More details on their contributions to this work may be found in the "Statement of Contributions of Authors" (q.v.). I also wish to thank my thesis reviewers and defence committee -- Jocelyn Faubert, Fred Kingdom, Dan Levitan, Gillian O'Driscoll, and the external reviewer, -- for thoughtful and useful comments on this work.

My tremendous gratitude goes to my partner Nicole Varshney for her unflagging support throughout this endeavor. I could not have done it without her.

I also wish to express my thanks to Dr. Olga Overbury, a peerless mentor who got me on the path.

Last but not least I must thank my parents, Roger and Margaret Collin, for everything they've done for me. They are the best.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

There has recently been much interest in how limitations in spatial frequency range affect face and object perception. This work has mainly focussed on determining which bands of frequencies are most useful for visual recognition. However, a fundamental question not yet addressed is how spatial frequency overlap (i.e., the range of spatial frequencies shared by two images) affects complex image recognition. Aside from the basic theoretical interest this question holds, it also bears on research about effects of display format (e.g., line-drawings, Mooney faces, etc.) and studies examining the nature of mnemonic representations of faces and objects. Examining the effects of spatial frequency overlap on face and object recognition is the main goal of this thesis.

A second question that is examined concerns the effect of calibration of stimuli on recognition of spatially filtered images. Past studies using non-calibrated presentation methods have inadvertently introduced aberrant frequency content to their stimuli. The effect this has on recognition performance has not been examined, leading to doubts about the comparability of older and newer studies. Examining the impact of calibration on recognition is an ancillary goal of this dissertation.

Seven experiments examining the above questions are reported here. Results suggest that spatial frequency overlap had a strong effect on face recognition and a lesser effect on object recognition. Indeed, contrary to much previous research it was found that the band of frequencies occupied by a face image had little effect on recognition, but that small variations in overlap had significant effects. This suggests that the overlap factor is important in understanding various phenomena in visual recognition. Overlap effects likely contribute to the apparent superiority of certain spatial bands for different recognition tasks, and to the inferiority of line drawings in face recognition. Results concerning the mnemonic representation of faces and objects suggest that these are both encoded in a format that retains spatial frequency information, and do not support certain proposed fundamental differences in how these two stimulus classes are stored. Data on calibration generally shows non-calibration having little impact on visual recognition, suggesting moderate confidence in results of older studies. [348 words]

RÉSUMÉ

Il y a eu dernièrment beaucoup d'intérêt sur les effets qu'a la gamme de fréquences spatiales sur la perception des objets et des visages. Ce travail c'est surtout concentré sur la question des gammes de fréquences les plus utiles pour la reconaissance visuelle. Une question fondamentale n'a pas encore été addressée. Celle-ci concerne l'effet qu'a le chevauchement des gammes de fréquences spatiales de deux images sur la reconaissance. À part son intérêt théorique. cette question a ausie des implications pour les études examinant les formats représentationels des stimuli (e.g., dessins, visages Mooney, etc.) ou la nature des représentations mnémoniques des objets et des visages. Cette thése examine principalement les effets de chevauchement des gammes de fréquences spatiales sur la reconnaissance.

Une deuxième question examinée est reliée à l'effet qu'a la calibration des stimuli sur la reconaissance des images filtrées spatialement. Les études antérieures qui ont utilisées des stimuli non-calibrées ont introduit, par inadvertence, des fréquences spatiales anormales à leurs images, ce qui aurait pu avoir un effet sur la reconaissance. Le fait que cette question n'a jamais été examiné met en doute l'équivalence des résultats de ces études par rapport à ceux d'études plus récentes.

Sept expériences sont discutées ici. Les résultats suggèrent que, contrairement à ce qui avait été proposé précédemment, la gamme de fréquences spatiales d'une image a peu d'effet sur sa reconnaissance, et que le chevauchement de ces fréquences est la determinante principale de la reconnaissance. Ceci suggère que le chevauchement des fréquences spatiales pourrait expliquer certains phénomènes de reconaissance visuelle. Quant aux représentations mnémoniques des visages et des objets, les résultats suggèrent que ces deux types d'image sont retenues en mémoire dans un format qui contient de l'information sur les fréquences spatiales. Ces résultats ne sont pas compatibles avec l'idée qu'il existe des différences fondamentales entre les représentations de ces deux types de stimuli. Finalement, la calibration a généralement eu peu d'impact sur la reconnaissance, suggérant que les études qui ont utilisées des stimuli non-calibrés sont généralement fiables. [332 mots]

STATEMENT OF CONTRIBUTIONS OF AUTHORS

The data from Experiments 1 to 5 of this thesis were previously published in the Journal of Experimental Psychology: Human Perception & Performance, with Liu, Collin, Rainville and Chaudhuri as authors (in that order). For these experiments, Collin made important contributions to the initial hypothesis formulation and development, as well as the design of the experiments. He also provided major contributions to the execution of the studies and analysis of the data. The text of the reported experiments in the dissertation was solely written by Collin. Liu was responsible for the initial hypothesis formulation of these experiments, although Collin helped develop and expand these ideas. Liu made important contributions to the design of the studies, as well as their execution. Rainville aided in the creating the computer programs which were used to present stimuli to subjects in these experiments. Chaudhuri aided in a general and supervisory manner on these studies, providing feedback on design, data analysis and manuscript preparation.

The data from Experiments 6 and 7 of this thesis are presently under submission at the Journal of Experimental Psychology: Human Perception and Performance, with Collin, Liu, and Chaudhuri as authors (in that order). Collin was responsible for the hypothesis formulation and design of these studies. He was primarily responsible for their execution and the analysis of the data as well. The text of the reported experiments in the dissertation was solely written by Collin. Liu made useful contributions to the hypothesis formulation and design of these experiments, and aided in their execution. Troje aided in these experiments by providing the 3D head models which were used as stimuli. Chaudhuri aided in a general and supervisory manner on these studies, providing feedback on design, data analysis and manuscript preparation.

Both of the manuscripts from which the data are drawn are referenced on page 54 of the thesis, and full bibliographic information is given in the references section.

CHAPTER 1: INTRODUCTION

This dissertation examines the effects of varying spatial frequency overlap on face and object recognition. This is done in an attempt to provide a better understanding of how we see and how we remember what we see. Spatial frequency overlap simply refers to the degree to which two images share spatial frequency bandwidth (a more extensive description is given in a later section). This factor has been little examined in the past and may help in understanding the sometimes contradictory body of findings on spatial frequency and complex image recognition.

A second area which is investigated is the effect of calibration on recognition of spatially filtered images. A number of researchers (Metha, Vingrys & Badcock, 1993; Olds, Cowin & Jolicoeur, 1999; Pelli & Zhang, 1991; Tyler & McBride, 1997) have noted that failing to linearize the relationship between pixel values -- the input for filtering algorithms -- and luminance values -- the output of stimulus presentation apparatus -- results in the inclusion of aberrant frequency elements in a filtered image. However, while there have been a number of articles on how to linearize this relationship, there have been none investigating exactly how non-calibration might affect recognition. This is important because many early studies (e.g., Fiorentini, Maffei, & Sandini, 1983; Millward and O'Toole, 1986) used slides or other stimuli which produced a non-linear luminance function. If non-calibration has strong effects, then this would make interpretation of these early studies relative to more modern ones quite difficult.

The dissertation begins with a brief history of the field of face and object perception in general. This provides a broad context for the rest of the research presented here. The historical review concentrates on the controversy over whether faces are treated in a way that is qualitatively different from objects, and includes a discussion of recent challenges to the idea of face "specialness." The subsection concludes with a summary rationale for studying spatial frequency overlap effects on face recognition.

Following the historical outline is a detailed literature review. This examines past research on spatial vision as it applies to complex image perception, especially concentrating on face recognition. The literature is divided into four major segments: 1) Research on coarse-quantization (also called pixelization or block averaging), 2) Studies of recognition across display form, such as work on the recognizability of line drawings, 3) Research on which bands of spatial frequencies are most useful for various visual tasks, and 4) Studies of how spatial frequency information is used in storing face and object representations in memory. The literature review closes with a detailed rationale for studying spatial frequency overlap. This also includes a brief rationale concerning the effects of calibration and a review of what little previous work has been done concerning this question. This concludes the first chapter.

The second chapter of the dissertation presents seven novel experiments on spatial frequency overlap as it affects face and object recognition (there are in fact 12 experiments, but 10 of them are paired into the first five for the sake of organization). These use a variety of methods, including learn/test recognition, simultaneous matching and sequential matching. All show a strong effect of overlap, suggesting that this factor should be taken into account when analyzing a variety of higher-order visual phenomena.

The third and final chapter of this document is the general discussion. This is an attempt to bring together the findings of past research with those presented here to provide a better overall understanding of visual recognition. This chapter re-iterates the form of the rationale given in the introduction, examining each of the questions it poses in a separate subsection. The discussion closes with suggestions for further research and a summary of basic conclusions which can be drawn from this work.

A Brief History of Face Perception Research

Research in face perception has enjoyed a steady rise in popularity over the past few decades. Whereas there were fewer than 30 scholarly articles published on this topic during the 1950's (Ellis, 1986), there were over 600 published articles in the 1970s. More recent decades have seen thousands of articles on face perception published, as well as several books (Bruce, 1988; Bruce & Young, 1998; Bruce, Cowey, Ellis, & Perrett, 1992; Bruyer, 1983; Ellis, Jeeves, Newcombe, & Young, 1986; Young, 1998). The reasons for this rising interest are many and complex, but likely have much to do with the intuitively compelling nature of human faces. Faces are a conduit of communication between people that arguably rivals spoken language in importance. They are also the most important and central means for assessing individual identity on an everyday basis. Given the great social relevance of face perception, what is perhaps surprising is the lack of attention it received prior to the 1970s. Though Darwin (1872) wrote extensively about facial expression and several early psychologists touched on related topics (see Goldstein, 1983), systematic research has only been undertaken in earnest in the last 25 years or so.

One reason for the relative neglect of face perception as an area of study in the past may have been that faces were assumed to be recognized by the same mechanisms underlying general visual perception. As such, face recognition did not warrant study in isolation. Indeed, the recent surge of interest in this topic seems largely fueled by the intriguing suggestion that faces are processed by the human visual system in a manner qualitatively different from other objects. Teuber (1968) was the first modern scientist to propose this possibility, and a number of researchers have since provided evidence to support the idea. *Prima facie* the suggestion that humans might have a "face processing module" seems quite plausible, for it would have been of paramount importance for our primate ancestors to correctly identify each other and to accurately assess each other's emotional states (Anderson, 1994). As we are a social species heavily dependent on the sense of vision, it seems reasonable that a powerful and efficient means of accomplishing these tasks using visual cues would have evolved in us, just as similar scent-based mechanisms have evolved in species who rely on that sensory modality. Supporting the idea of face "specialness" is a great deal of evidence suggesting a qualitative difference between how faces, as opposed to objects in general, are visually processed. As we will see in this review, however, the underlying source of this apparent difference remains contentious.

Yin (1969, 1970) was the first to report behavioural evidence for qualitative differences between face and object processing. He found that upside-down faces were disproportionately difficult for subjects to recognize as compared to upside-down objects. Since then many studies have replicated his findings in humans and other higher primates (e.g., Bartlett & Searcy, 1993; Enns & Shore, 1997; Farah, Tanaka & Drain, 1995; Parr, Dove, & Hopkins, 1998; Phelps & Roberts, 1994; Pullan & Rhodes, 1996; Tomonaga, 1994; Vermeire & Hamilton, 1998; Wright & Roberts, 1996; see Valentine, 1988 for a review of earlier work) and a number of other phenomena have been uncovered that seem to point to differences in the way the visual system treats faces vs. objects (for reviews see Bruce, Cowey, Ellis, & Perrett, 1992; Bruce & Humphreys, 1994; Bruce & Young, 1998; Farah, 1996; Tovee, 1998; Tovee & Cohen-Tovee, 1993; Young, 1998). For example, face recognition is disproportionately affected by contrast reversal, also known as photographic negation or brightness reversal (Anstis, 1992; Bruce & Langton, 1994; Galper, 1970; Galper & Hochberg, 1971; Gauthier, Williams, Tarr, & Tanaka,1998; Hayes, 1988; Hayes, Morrone, & Burr, 1986; Johnston, Hill, & Carman, 1992; Kemp, McManus, & Pigott, 1990; Kemp, Pike, White, & Musselman, 1996; Liu & Chaudhuri, 1997, 1998; Luria & Strauss, 1978; Phillips, 1972). Also, variations in lighting direction (Braje, Kersten, Tarr & Troje, 1998; Bruce, 1998; Enns & Shore, 1997; Hill & Bruce, 1993, 1996; Johnston et al., 1992) and rotation in depth (Bruce, Valentine & Baddeley, 1987; Davies, et al., 1978; Krouse, 1981; Logie et al., 1987; Patterson & Baddeley, 1977; Schyns & Bulthoff, 1994; Troje & Bulthoff, 1994; Wogalter & Laughery, 1987) have greater effects on face recognition. Lastly, it has been suggested that faces are processed holistically while objects are processed in a part-based manner (Biederman and Kalocsai, 1997; Bradshaw & Wallace, 1971; Tanaka & Farah, 1993; also see Bruce, 1988 for a review).

This behavioural evidence has been supported by a number of neurological studies that have suggested that face processing is supported by its own brain module, located in the middle fusiform gyrus. These include brain imaging studies (Aguirre, Singh, & D'Esposito, 1999; Haxby et al., 1999; McCarthy, Puce, Gore, & Allison, 1997) as well as clinical and electrophysiological studies (see Farah, 1996 for a review). The McCarthy et al. (1997) study is typical of this body of research. In this study, subjects were shown face and flower images that were either scrambled or unscrambled. The scrambled condition was necessary to control for or "subtract out" general visual recognition areas. It was assumed that these sorts of images would activate basic visual areas as well as feature and shape detecting modules, but not higher centers responsible for face or object recognition. McCarthy et al.'s (1997) results support the notion of a face specific area. They found that when faces were viewed amidst unscrambled objects there was bilateral activation in the fusiform gyrus, but when faces were viewed amist scrambled objects, there was focal activation in the right fusiform gyrus. These results are compatible with a model in which the fusiform gyri are the seat of a general object recognition system, while an area of the right fusiform gyrus is specific to face recognition.

The special status of faces in cognitive and visual systems has recently been challenged by the work of Gauthier and colleagues (Gauthier, 1999; Gauthier, Behrmann, & Tarr, 1999; Gauthier & Tarr, 1997; Gauthier, Tarr, Anderson, Skudlarski & Gore, 1999; Gauthier, Williams, Tarr & Tanaka, 1998; Tanaka & Gauthier, 1997). These researchers argue that many of the phenomena that are held to distinguish faces from objects are in fact the result of differences in subject expertise with the two sets of stimuli. Gauthier (1999) points out that whereas normal human beings have great expertise in recognizing faces, they do not have similar expertise with other sorts of objects. She predicts that if subjects are well trained to recognize and discriminate an object class with similar physical characteristics to faces, then it will elicit the same inversion, lighting, and contrast reversal phenomena that faces do. If this is the case, it argues against studying face perception in isolation from object perception. Instead, the appropriate course would be to include an expertise factor in models of general visual recognition, which would thus be able to predict face recognition performance without resort to less parsimonious special mechanisms.

In several studies, Gauthier and her colleagues had participants train extensively to recognize a class of artificial objects known as "greebles" and then tested them to see if they showed the same effects with these stimuli as they did with face images. For example, Gauthier and Tarr (1997) found that subjects trained to be "greeble experts" did indeed exhibit the kinds of holistic effects seen in face recognition, though in terms of accuracy they did not show a similar inversion effect, nor an effect of brightness reversal. Gauthier and Tarr (1997) found some evidence for an inversion effect in expert greeble-viewers, but only in the form of a reaction time (RT) difference, not an accuracy deficit as seen with face stimuli. Also, the RT deficit is only seen on the second of two blocks of test trials. The authors interpret this in terms of non-expert subjects adapting to the new inverted stimuli faster than expert subjects. However, this is not the pattern of behaviour one would predict with face stimuli, which would be expected to show the inversion effect on the initial testing trial as well as any subsequent ones. That is, normal subjects -- who presumably have been training since infancy in upright face recognition -- show an inversion effect any time they are shown inverted faces, including initial presentation. They do not merely adapt to them more slowly. Finally, the inverted greebles were found to be more difficult for both novices and experts, indicating that there is something innately more difficult in dealing with upside-down greebles than right-side up ones. This should not have been the case if the objects being presented were truly novel and the novices had no prior expectations of them.

Gauthier et al. (1999) also challenge previous findings of a face specific area of the infero-temporal cortex in or around the middle fusiform gyrus (Puce, Allison, Asgari, Gore & McCarthy, 1996; Damasio, Damasio, & Van Hoesen, 1982; Damasio, Tranel, & Damasio, 1990; Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Spencer, Spencer, & McCarthy, 1997; Puce, Allison, Gore, & McCarthy, 1995). In their study, they examined both novices and greeble experts using functional magnetic resonance imaging (fMRI) and found that the middle fusiform gyrus was activated in the latter group when viewing greebles, just as it was with normal observers viewing faces. Gauthier (1999) also reports similar results with bird and car experts when they viewed their stimuli of expertise. She and her colleagues interpret these findings as indicating that the middle fusiform gyrus is not specifically dedicated to faces, but rather to the recognition of any relatively homogeneous class of stimuli with which one has much experience. In this interpretation, the module is devoted to some process that develops as one gains aptitude in discriminating small differences in a visually similar class of objects. The nature of this process is difficult to determine, but may involve higher-order configural processing (Gauthier, 1999).

Several criticisms have been leveled at Gauthier's work. For instance, Biederman and Kalocsai (1997) have noted that greebles are somewhat face-like in their arrangement. They are composed of two rounded masses analogous to a body and a head. Three spikes emerge from the "head" in positions appropriate for ears and a nose, while a single spike emerges low on the "body", which may be seen as a penis. Biederman and Kalocsai argue that such feature arrangements may simply be eliciting face-like effects because of their similarity to faces. That is, greeble experts may not be learning to identify greebles *per se*, but rather learning to see greebles as faces. Greebles might also be seen as heads with a nose -- where Biederman and Kalocsai see a penis -- and an unusual hairstyle or horns. This interpretation is even more amenable to processing by a face-specific module. In response to these criticisms, Gauthier and colleagues are presently developing a new class of novel objects called "Yufos" (Gauthier, 1999). These have a distinctly non-face-like appearance, but no results using these stimuli have yet been published.

Even if Gauthier and colleagues' conclusions are correct, however, this does not necessarily mean that there is no specific mechanism for face recognition, nor does it argue conclusively against the role of the middle fusiform gyrus as the seat of this mechanism. Indeed, given that this cortical area is devoted to higher-order configural computations or some other process necessary for the discrimination of visually homogeneous objects, it may nonetheless be devoted to face recognition in a phylogenetic sense. That is, the area may have evolved its capacities due to evolutionary pressures related to face recognition only to be recruited, with extensive training, for other tasks. A similar situation has already been recorded concerning another area of the brain, specifically the left inferior frontal lobe, containing Broca's area. Although it is widely accepted that this area is devoted to language processing, it activates in expert musicians when they are exposed to tunes (e.g., Hugdahl, et al., 1999; Penhune, Zatorre, & Evans, 1998). Clearly the area did not evolve for music perception, but due to the plasticity of cortical function it can be recruited for this task given sufficient training. Similarly, it seems unlikely that the capacity to discriminate subtly different forms within object classes would have evolved for the purpose of identifying cars or birds. Indeed, it is difficult to think of any class of natural objects other than faces whose subtle individual discrimination is of adaptive significance to primates under natural conditions. Other animals and plants need only be identified by species to allow appropriate responses, individual discrimination is unnecessary.

Based on the above argument, it seems highly plausible that we possess an innate ability to make subtle intra-class discriminations and that this evolved for the purpose of face perception. The most likely candidate for a cortical center subtending this ability remains the middle fusiform gyrus. The fact that this area can, given sufficient practice, be recruited for other tasks does not counter this assertion.

In summary, Gauthier and colleagues have convincingly shown that higher-order configural processing is involved in face recognition and that this type of processing can be recruited for other visually homogeneous object classes given high levels of training. But their work has not shown inversion effects nor contrast reversal effects on accuracy in greeble experts. To date, the photographic negative effect, depth rotation effects and a number of other phenomena continue to qualitatively distinguish face recognition from the recognition of other object classes. Thus, the idea of studying face perception in isolation from general object recognition remains valid.

If one accepts that faces are indeed "special", that they are processed in some way qualitatively different from other objects, one is then faced with the question of what the nature of this difference might be. One possibility is that face and object recognition make primary use of visual information at different spatial scales. It may be that the two processes make primary use of different spatial bands, or it may be that the stored representations retain different bands. Another possibility that has been suggested is that object representation might rely on higher-order features that are spatial-frequency-free whereas face representations store matrices of raw low-level filter outputs (Biederman & Kalocsai, 1997; Kalocsai & Biederman, 1998).

Along these lines, several researchers have proposed that whereas objects are recognized primarily based on high spatial frequency edge information, faces are recognized by surface properties that incorporate both coarse and fine spatial frequencies (see Bruce & Humphreys, 1994, for a review). Evidence for this comes from studies of recognition of images in different display formats (e.g., line drawings, photographs and two-tone images), which are held to preferentially preserve different spatial frequency bands. Full bandwidth photographs of faces are better recognized than either line drawings or two-tone images, which are thought to preferentially preserve high- and low-frequency information respectively (Bruce, Hanna, Dench, Healey, & Burton, 1992; Davies, et al., 1978). However, for object recognition, line drawings are about as effective as photographs (Biederman & Ju, 1988).

Object recognition may also be more robust to variations in spatial frequency content in general than face recognition. Low-pass or high-pass face images containing a narrow band of frequencies are more difficult to recognize than those containing a broader band of frequencies, which in turn are more difficult to recognize than full-bandwidth images (Bachmann, 1991; Costen, Parker, & Craw, 1994, 1996; Fiorentini, Maffei, &

Sandini, 1983; Harmon, 1973; Millward, & O'Toole, 1986; Parker, Lishman, & Hughes, 1996). But studies on naming or matching of complementary images in the Fourier domain show that object recognition displays relative invariance to spatial frequency content (Biederman & Kalocsai, 1997; Kalocsai & Biederman, 1998).

The above shows that researchers consider knowledge about how spatial scale affects visual recognition to be important, and that they feel it may reveal interesting things about how face and object perception differ. Despite this, there are a number of questions related to spatial scale and visual recognition which have not been addressed. One of these concerns the effects of varying spatial frequency overlap. Although a few studies have touched on this factor (Biederman & Kalocsai, 1997; Kalocsai & Biederman, 1998; Millward & O'Toole, 1986), there have been no systematic attempts to evaluate the effects of it on recognition.

Instead, studies on the effects of spatial frequency variations in visual recognition have primarily concentrated on determining what ranges of spatial frequencies are most useful for the tasks. In object recognition, this initially took the form of a debate as to whether low or high spatial frequencies were of paramount import (e.g., Ginsburg, 1978, 1980; vs. Fiorentini, Maffei & Sandini, 1983), an issue that is still being examined (Parker et al., 1996). In face recognition, a series of studies points to a "critical band" of middle object frequencies as most important (Bachmann, 1991; Costen, Parker, & Craw, 1994, 1996; Fiorentini et al., 1983; Harmon, 1973; Hayes, et al., 1986; Peli, Lee, Trempe, & Buzney, 1994; Gold, Bennett, & Sekuler, 1999; Nasanen, 1999).

Although determining the relative usefulness of different spatial bands provides us with some useful data for making predictions about the effects of spatial frequency on human recognition performance, it does not provide sufficient information for modeling human behaviour under certain circumstances. To predict performance at some tasks, it is necessary to know something about how the degree of spatial frequency overlap affects recognition. It is clear, for instance, that matching a high-pass image to a low-pass one is more difficult than matching either of these to an unfiltered original. Part of the reason for this is likely that the high-pass and low-pass images share little or none of the spatial frequency spectrum. To the extent that different display formats (line drawings, photographs, two tone images) present different spatial frequency bands, information on overlap effects will also be important in determining how humans can match across variations in such representations.

Another motivation for examining the overlap factor is that such investigations can give us insights into how faces and objects are represented in the brain. Biederman and Kalocsai (1997; Kalocsai & Biederman, 1998) have proposed a model wherein face representations retain spatial frequency information but object representations do not. That is, they hold that faces are mnemonically stored in the form of vectors in a multi-dimensional metric space, with the dimensions representing the activation levels of an array of early filters. As such, these representations retain information about the spatial frequency of different detected features. In contrast, objects in this model are thought to be represented as Geon Structural Descriptions (GSDs) which are qualitative descriptions consisting of a number of volumetric primitives and terms describing their aspect ratios, their relations to one another, and so on. GSDs do not retain spatial frequency information and can be activated by a broad range of early filter activation patterns, thus providing objects with greater robustness to spatial frequency variations. If this is the case one would expect object images to show greater tolerance for low levels of spatial frequency overlap than face images. Biederman and Kalocsai (1997; Kalocsai & Biederman, 1998) report results that suggest exactly this, though their paradigm examines only the case of no overlap vs. complete overlap. An examination of more gradual changes in overlap might therefore be more informative as to how this difference comes about.

This dissertation examines the effects of spatial frequency overlap on face and object recognition. This is done with the goal of better understanding the issues discussed above, as well as several other questions in the general area of visual recognition. As a first step. past research on spatial scale as it applies to complex image perception is reviewed in detail in the next subsection. This concludes with a detailed rationale for the study of spatial frequency overlap, which includes a set of specific question which need to be addressed.

A Note On Units

In studies of spatial vision, spatial frequency is typically measured in units of cycles per degree (c/d). However. it is widely agreed in the literature on recognition that a better unit for examining complex images is cycles per object or cycles per face. Research has shown (e.g., Parish & Sperling, 1991) that varying the viewing distance -- and thus the absolute spatial frequency content -- of stimulus images does not affect complex image recognition over a wide range. For this reason, relative units, using stimulus dimensions as a base, are generally used in the literature. This dissertation will follow this convention, using "cycles per object" (c/o) to describe research on both face and object recognition wherever possible. Unless otherwise noted, this unit refers to the number of sinusoidal cycles that cross the width of the actual face or object stimulus (as opposed to the complete stimulus image, which may incorporate an irrelevant background of varying size). In some cases, the information necessary to calculate c/o is not given and estimates based on limited information are used. These are clearly labeled as such. Cycles per degree (c/d) will also be used on occasion, and refer as usual to the number of cycles that cross an area subtending one degree of visual arc.

Literature Review

To date, four major lines of research have dominated the literature on spatial frequency in complex image recognition. Early work concerned the effects of pixelizing faces and the fact that this caused recognition to decline until the image was blurred or viewed from a distance. Two basic theories as to the reason for this phenomenon were advanced, one which held that pixelization added high-frequency noise to an image, thus masking the contents of the original (e.g., Harmon & Julesz, 1973), and another which held that the phenomenon arose from a spatial mislocation of landmarks in perceptual organization (e.g., Morrone et al., 1983; Morrone & Burr, 1994).

At the same time, another group of researchers were examining the effects of varying mode of representation, investigating the efficacy of representations such as line drawings and two-tone images relative to photographs. This work, originally very practical and applied in nature, branched off in a more theoretical direction concerned with which bands of spatial frequencies were most useful for recognition. Early research on this question debated the efficacy of high vs. low frequencies in object recognition. Later research examined face recognition and suggested the importance of a middle band between 8 and 16 c/o.

Recently, the issue of which bands of frequencies are most useful to the recognition process has spawned a fourth line of research, this one concerning how spatial frequency information is used in representing faces and objects in memory. This has been most directly examined by Biederman and Kalocsai (1997; Kalocsai & Biederman, 1998), who propose a model in which face representations retain information on spatial frequency while object representations discard it (see also Biederman, 1993, for the apparent genesis of the idea). This line of inquiry represents an interesting new direction for studies in spatial vision, bringing issues of spatial frequency into the cognitive domain.

The following literature review will examine the four issues mentioned above, treating each in detail. The review concludes with the observation that attention to the factor of spatial frequency overlap may be of use in providing a more complete understanding of how we recognize what we see. Note that although these four research directions are presented as developments of one another, they should not be taken to have occurred in any definite chronological order. As will be seen, they have generally paralleled each other throughout the last half of the 20th century. Only the relative degree of interest in each has changed.

Recognizing Pixelized Images: Critical Band Masking vs. Spatial Mislocation

Harmon and Julesz (Harmon, 1971; Harmon & Julesz, 1973) were the first researchers to examine spatial frequency effects using complex stimuli. They had subjects attempt to recognize images that had been pixelized. That is, the image was divided into a grid of square cells and all pixels in each cell were set to the mean of their original set of values. This procedure is sometimes called pixelization, coarse-quantization or block-averaging. Harmon and Julesz (Harmon, 1971; Harmon & Julesz, 1973) noticed that recognition accuracy improved if the pixelized images were blurred or viewed from a distance. They hypothesized that block averaging, in addition to reducing the resolution of the image and effectively low-passing it, also impaired recognition by producing high-frequency spatial noise that masked the original image. Block averaging produces spurious spatial frequency elements whose amplitude drops off linearly as frequency rises above the fundamental frequency of the quantization grid (Harmon, 1971). Harmon and Julesz (Harmon, 1971; Harmon & Julesz, 1973) proposed that the noise adjacent in frequency to the signal in a complex image would produce the greatest deficit in recognition, in a manner analogous to critical band masking seen in audition (e.g., Scharf, 1971; Moore, 1995) and basic vision (e.g., Blakemore & Campbell, 1969; Campbell & Robson, 1968; Morgan & Watt, 1984). To examine this hypothesis, they repeated their experiments using blocked images that had been band-stop filtered to remove elements either immediately above the frequency of the quantization grid or to remove elements more remote in spatial frequency. Leaving in elements at frequencies adjacent to the image frequencies indeed caused greater impairment than elements that were more distal in frequency. For instance, in a picture block-averaged at 10 c/o (i.e., 20 pixels per image), leaving in the noise between 10 and 40 c/o width had great effect on recognition accuracy while noise between 40 and 70 c/o had little effect.

Morrone, Burr & Ross (1983) question the interpretation offered by Harmon (1971) and Julesz (Harmon & Julesz, 1973), noting that the energy of the high frequency components introduced by coarse-quantizing an image is too low to mask the low frequency components of the original picture, which are generally high in energy in natural images. Morrone et al. (1983) show that introducing noise at oblique spatial orientations renders block-averaged images easier to perceive. This is incompatible with the masking hypothesis, as according to this view adding more high-frequency noise should only produce greater difficulties in recognition.

One possible explanation of Morrone et al.'s (1983) results is that the oblique noise they added may have masked the noise already introduced by block averaging. In essence, the added energy may have "canceled out" that which was introduced by coarse-quantization. However, the noise added by Morrone et al. (1983) was dissimilar in orientation to that introduced by block averaging, occupying only those areas within 22.5° of the oblique in each quadrant of the Fourier domain (i.e., orientations from 22.5 to 67.5°, 112.5 to 157.5°, 202.5 to 247.5, and 292.5 to 337.5). Morrone et al. (1983) argue that this rules out the possibility of low-level masking by the oblique noise of the horizontal/vertical noise because elements of dissimilar orientation do not mask each other under most circumstances.

The explanation favored by Morrone et al. (1983) is that the grid structure introduced by block averaging interferes with higher-order perceptual organization mechanisms designed to perform figure-ground segregation and overall structural analysis of images, and that adding the oblique noise removes this effect. That is, when a picture is block-averaged, a high-frequency structure is introduced along the blocks' edges. This is made more salient than the original structure of the face, producing a form of high-level shape masking. Adding oblique noise at the same frequencies as the blocks' noise causes a breakdown of the block structure, allowing low-frequency higher-order mechanisms to detect and extract the structure of the original face image.

Durgin and Proffitt (1993) report another experiment that challenges Harmon and Julesz's model. They found that super-imposing a grid of black lines over the quantization grid produced superior recognition, with the face appearing to be "behind" the grid. They interpret this finding as being counter to the noise hypothesis, claiming that they are in effect adding more quantization noise and yet improving recognition. But the results are difficult to interpret, as adding solid black lines to the image introduces spatial elements at all frequencies, not just high ones. More importantly, the black lines differ strongly in luminance from the other elements in the picture and could thus be luminance-masking the original noise that is at adjacent frequencies to the image elements.

In an experiment similar to Durgin and Proffitt's (1993), Morrone and Burr (1994) found that reversing the contrast of the edges produced by the blocking operation had similar effects. Again, performance improved and again face images appeared to be "behind" the contrast-reversed grid. Although this manipulation eliminates concerns about luminance differences, it nonetheless produces new sharp edges in the image that introduce harmonic elements across a broad range of spatial frequencies. A later experiment by the same researchers (Morrone & Burr, 1997) varied the phase of the spurious harmonics introduced by block averaging. They found that with large phase shifts, subjects improved in recognition, reporting as with the earlier experiments (Morrone & Burr, 1994; Durgin & Proffitt, 1993) that the low frequency elements appeared "behind" the phase-shifted quantization grid. This assuages the methodological concerns regarding similar earlier experiments and makes a strong argument against the idea that noise is solely responsible for the effects of quantization. Taken together, this group of experiments (Durgin & Proffitt, 1993; Morrone & Burr, 1983; Morrone & Burr, 1994; Morrone & Burr, 1997) supports the involvement of higher-order perceptual organization mechanisms in the phenomenon.

Costen, Parker and Craw (1994, 1996) also examined the effects of low-pass filtering vs. coarse-quantizing face images. They found that coarse quantization causes a more rapid decline in performance as the cut-off frequency is lowered. They attribute this to a mixture of noise effects and mislocation of features in the spatial domain, in essence agreeing with the idea that noise effects cannot be the sole explanation for pixelization phenomena. It should be noted, however, that their studies were not designed to test either theory explicitly, so strong conclusions cannot be reached from their data as to what the relative contributions of noise and higher-order effects might be.

Uttal, Baruch, and Allen (1995a; 1995b; 1997) have performed a series of studies that examines the effect originally reported by Harmon and Julesz (1973). In the first (Uttal et al., 1995a), they examined the ability of subjects to discriminate small silhouettes of airplanes. They found that filtering such a block-averaged stimulus did not improve performance, but rather degraded it further. This result is the qualitative opposite of that found by Harmon and Julesz (1973). They argued from this that task differences exist in this phenomenon, a position similar to one earlier proposed by Sergent (1986). A second, closely-related study (Uttal et al., 1995b) used the same stimuli but in a recognition paradigm. Here they found results in agreement with Harmon and Julesz (1973), supporting the hypothesis that different tasks use spatial frequency information differently. Oddly, however, they found the same results whether they first block-averaged the stimuli and then filtered them or first filtered and then block averaged them. As the latter manipulation should introduce noise elements in the same way that blocking alone does, this result is incompatible with an explanation of the blocking/filtering phenomenon, which relies purely on high-frequency noise effects.

In their most recent study, Uttal et al. (1997) used a recognition paradigm and changed the stimuli to faces of two different sizes. With large faces (about 6° across) they find a pattern compatible with the noise hypothesis: Blocking then filtering creates superior recognition to blocking only, but filtering then blocking does not. However, with small face stimuli (about 1° across) they find that either order of image degradation produces similar results. That is, whether one quantizes an image and then filters it or filters it and then quantizes it, subject accuracy is better than with the quantized image alone. The researchers suggest that it is not the size of the stimuli *per se* that is causing differences in the results, but rather a complex interplay between type of task, size of the stimuli, size of the blocks, and the filter parameters. Unfortunately, they do not detail their position any more than this.

To summarize the above literature, it seems that Harmon and Julesz's (1973) original explanation for the improvement in recognizing coarse-quantized images cannot be the entire explanation. Although noise effects may certainly be involved, there is also a great deal of evidence for higher-order mechanisms taking part (e.g., Durgin & Proffitt, 1993; Morrone et al., 1983; Morrone & Burr, 1994) and some suggestion that the phenomenon works differently at different spatial scales (Uttal et al., 1997). In retrospect, this seems entirely likely, as the "noise" introduced by block averaging is structured and therefore something of ecological relevance to the visual system. Certainly a grid of squares is something we would expect the visual system to extract from a scene, so what the block averaged image presents is not so much a signal obscured by noise, but two competing signals, both of which have higher-order structure, and both of which cover a wide range of spatial frequencies. A variety of manipulations can reduce the saliency of one signal or the other, eliminating this competition and allowing a single clearer image to emerge (or alternately, disentangling the two signals so they can both be perceived clearly). Unfortunately, this explanation is not entirely satisfactory, as the characteristics of higher-order perceptual organization have proven difficult to define formally and remain poorly understood relative to lower-level mechanisms. For instance, while it intuitively clear what the two competing signals are in a pixelized image, it is difficult to formally define which elements belong to each signal.

Research on coarse-quantization effects was only one antecedent of the current interest in spatial frequency effects on visual recognition. Another important precursor was work examining the effects of different display formats (e.g., photographs, line drawings, two-tone images, and so on) on face and object recognition. Initially this line of experimentation was concerned mainly with purely practical matters such as whether line drawings were effective means of identifying criminals to the public. But later researchers recognized that the issue was related to the spatial frequency content of the images and they manipulated display format to explore more fundamental questions.

Recognizing Images Across Different Display Formats.

A number of early studies examined the effectiveness of various modes of representation on face and object recognition. Mainly these were concerned with practical matters. For instance, Ryan and Schwartz (1956) investigated the effectiveness of line drawings vs. photographs in presenting complex visual information such as electrical diagrams and machine schematics. They found that photographs were identified more quickly than line drawings but that the latter produced a better understanding of the material being presented. Fraisse and Elkin (1963) examined the ability of subjects to recognize items represented as line drawings, shaded drawings, photographs or actual objects. Surprisingly, shaded drawings proved more recognizable than photographs in general, while outline drawings along with actual objects were hardest to recognize.

Davies, Ellis and Shepherd (1978) were the first to examine the question of display format with regards to face perception (In scientific terms, that is. The issue has of course been of interest to artists for considerably longer!). They investigated the effectiveness of outline drawings, detailed line drawings and photographs in a recognition paradigm. Their study was motivated by the question of which method of representation would be best for reconstructing faces of criminals, but it nonetheless provides some interesting hints as to how spatial frequency content affects recognition. In their first experiment, they tested subjects' recognition of faces of celebrities, gathering measures of recognition (the name associated with the stimulus) and familiarity (a rating on a 5-point scale). Both indicators suggested a strong advantage for photographs over detailed drawings, which were in turn superior to outline drawings. A second experiment along similar lines examined recognition of previously unknown faces in an old/new paradigm. Here subjects examined either detailed drawings or photographs in a learning session and were subsequently tested on photographs only in the testing session. Photographs continued to show a superiority, but this was much reduced as compared to the first experiment. These findings are opposite to those of previous studies (Fraise & Elkin, 1963; Ryan & Schwartz, 1956), finding that photographs produced better recognition than detailed drawings, which in turn were better than outlines.

The obvious explanation for the differences between Davies et al.'s (1978) findings and those of previous studies (Fraisse & Elkin , 1963; Ryan & Schwartz, 1956) lies in the types of stimuli they used. Davies used faces whereas the others used objects. It may be that different sorts of information are optimally useful for the purposes of face and object recognition. Recently, this idea has been reiterated by Biederman and Kalocsai (1997; Kalocsai & Biederman, 1998) who posit that whereas face representations retain a full range of spatial frequencies in their mental representations, objects discard spatial frequency in favor of a feature-based representation based mostly on edge locations.

Other evidence for the inadequacy of line drawings for face recognition was found by Kuehn and Jolicoeur (1993). Using a visual search task, they compared performance with line-drawn faces against performance with digitized face images. They found that both types of face elicited serial-search styles of visual search but that line drawings produced slower reactions overall. One could interpret this as evidence that low spatial frequencies are important to the face recognition task. However, this conclusion must be a guarded one as the full-bandwidth faces most likely contained a greater degree of information overall.

In addition to studies of line drawings, a number of studies have explored the recognition of two-tone images. These representations, also referred to as Mooney, lithographic or bi-quantized pictures, assign one of two values to each point in an image, usually a light white and a dark black. Hayes (1988) was the first to examine these sorts of faces with regards to what they might say about spatial frequency effects on recognition. He examined the recognition of both line drawings and two-tone images. To produce bi-quantized images Hayes (1988) performed a thresholding function on 256-level greyscale images at the mid-point of their luminance range. Line drawings were made by marking the points at which there was a transition from black to white in the bi-quantized image. The results of this experiment showed that two-tone images could be recognized quite well, whereas edge images produced near-chance performance. Contrast-reversed versions of the two-tone images also produced poor performance. This finding is unexpected because previous research on object-recognition showed that it was little affected by this manipulation (Sutherland, 1971). Hayes (1988) reasoned that the difference between his face stimuli and Sutherland's (1971) object stimuli might be that whereas the latter could be identified by their high spatial frequencies alone, the former required low-spatial frequencies for analysis. He supported this conclusion by showing nearly equivalent recognition performance for low-pass filtered two-tone and low-passed multi-tone images of faces (Hayes, 1988).

In another study of display format, Bruce et al. (1992) combined line drawing and two-tone images into "cartoons". They tested the ability of subjects to recognize each component of these cartoons (i.e., the line drawing alone and the two-tone image alone) as well as the combination. As with some of the previous studies, this one was directed at a practical issue, testing the efficacy of a 1-bit image transmission system for the hearing impaired. The line-drawings and bi-quantized images were produced by a system developed by Pearson and Robinson (1985) that created 1-bit images by combining an edge-detector with a thresholding function. This creates images that are subjectively quite recognizable, an impression that Bruce et al. (1992) attribute to the inclusion of the thresholded component, which provides some impression of the "mass" of the face. To test this assumption, they had subjects attempt to recognize images of famous persons that were rendered in one of four formats: Full greyscale images, line-drawings, thresholded images, and cartoons. Different groups of subjects were first asked to identify the photographs in order to determine a baseline for performance. About 90% of these images were recognized, with only small and non-significant differences between groups. Subjects were then asked to recognize the same images in the other three formats. Full cartoons (lines with thresholding) fared best, with a 93.3% relative accuracy. Images with only lines produced the worst performance, with a 67.2% relative accuracy. Images with only the threshold elements (bi-quantized images) were in between these two, with a relative accuracy of 77.4%. The fact that the full cartoons do better than either of the components is not surprising, as there is more information in the combination of the elements. But the fact that the threshold component was superior to the edge component is of interest, as it suggests that edges are not very useful for face recognition, while shading information is. This in turn suggests that lower spatial frequencies may be more important than higher ones.

Liu & Chaudhuri (1998) examined the recognition of two-tone vs. multi-tone faces in four conditions: Learn Positive/Test Positive (PP), Learn Positive/Test Negative (PN), Learn Negative/Test Negative (NN), and Learn Negative/Test Positive (NP). They found that two-tone images were recognized as well as multi-tone images in the PP condition, but that the two-tone images produced deficits in all other conditions. They attribute this pattern of findings to disruptions of different bands of spatial frequencies. Contrast reversal of an image is thought to interfere with the usefulness of low-frequencies because it creates impossible shadows and shading. Bi-quantization on the other hand, disrupts high spatial frequencies. If this is the case, and if low spatial frequencies are sufficient for face recognition, then simply bi-quantizing a face should not create insurmountable difficulties in its recognition. However, both bi-quantizing an image and contrast reversing it would hamper both the upper range of frequencies and the lower, making the task extremely difficult. This is exactly the pattern they found (Liu & Chaudhuri, 1998).

Biederman and his colleagues investigated the efficacy of line-drawings for object recognition in a number of studies (e.g., Biederman, 1987; Biederman & Ju, 1988; Biederman & Kalocsai, 1997; Kalocsai & Biederman, 1998). The study conducted by

Biederman and Ju (1988) is representative of this work. In this experiment, subjects were asked either to identify an object by naming it or to verify that presented names were congruous with presented pictures. They did this with either full-colour pictures of objects or with line drawings of the same. In all cases, accuracy and reaction time were virtually identical for the two types of representation. Biederman and Ju argue that this is due to the internal representations of object being edge based as opposed to surface-based, a finding that they take to support Biederman's (1987) Recognition by Components (RBC) model of object identification.

A summary of the research on cross-format recognition is difficult to give, as the wide variety of methods make it difficult to compare across studies. In general, however, it seems to be agreed upon that whereas line drawings are sufficient for object recognition, they are not enough to produce reliable face identification. It has been suggested that this implies a greater role for high frequencies in general object recognition, whereas middle or low frequencies are most important for face recognition. A more exact and quantitative conclusion is unfortunately not possible, as the image manipulations used in such studies have effects on spatial information content that are difficult to formally define. For instance, although line drawings are thought by many to present only high spatial frequencies to the observer, this is an incorrect assumption. Line drawings are in fact broad-band stimuli. Although a line drawing may include only high-frequencies from the original image upon which it was based, it also contains highly correlated information across the spatial spectrum. If a line drawing is derived from a full-bandwidth image by some kind of mathematical algorithm (e.g., high-passing and bi-quantizing), then it may be possible to determine the relationship between them in terms of spatial frequency content with some precision. But if the image is simply an artist's rendition of an object, as is often the case in such studies, then the exact relationship between the contents of the drawing and those of a normal full-bandwidth image are difficult to determine. Similarly, bi-quantizing an image, while disrupting the high spatial frequencies in the original, also adds in sharp edges highly correlated with the information in the original.

The exact consequences of these manipulations in terms of spatial frequency content are therefore hard to quantify and the nature of the information added by them is difficult to ascertain. Thus, although the manipulation of display format is an interesting issue in its own right, its use in exploring spatial frequency issues in visual recognition is somewhat limited. It is perhaps for this reason that most recent work on spatial frequency in recognition has focused more directly on the problem, using spatially filtered images or noise-masking techniques that allow more precision as to which spatial frequency elements are affected by the manipulation. It is to this line of work that I turn next.

Which Spatial Bands are Most Useful?

Perhaps the issue that has been of the greatest interest to researchers examining spatial frequency and visual recognition has been that of which spatial frequencies (if any) are most useful to the visual system for various tasks. Interest in this topic followed closely from Campbell and Robson's (1968) suggestion that the visual system processed scenes through a number of spatial frequency channels and continues to the present (e.g., Nasanen, 1999; Gold et al., 1999). This research is partly motivated by the possibility that different channels subserve different tasks, and partly by a debate over whether the processing speed advantage of low-frequency elements gives them a special role in higher visual processes. The latter issue has implications for models of face and object recognition, with some favoring a coarse-to-fine sequence and others emphasizing the importance of high-frequency information (Parker et al., 1996).

Harmon (1971) and Julesz (Harmon & Julesz, 1973), though primarily interested in determining an explanation for why filtering coarse-quantized images makes them more recognizable, also worked on determining the minimum number of picture elements that could drive recognition. Their experiments suggested that face recognition could be supported by a coarsely-quantized image containing as little as 16 x 16 pixels. Because their face images did not fill the entire grid, this corresponds (based on the examples provided in their papers) to about 4 c/o. Under these conditions, subjects could name faces of familiar people quite well. Accuracy was 48% (chance was 17%) when the quantization grid was randomly aligned with the original image, rising to 95% when the best of several grid alignments was used.

Although Harmon and Julesz were the first to spur interest in spatial frequency effects on recognition, it was Ginsburg (1978, 1980) who first explicitly hypothesized that a given band of spatial frequencies might be more useful for recognition than others. His experiments examined the effects of spatially filtering images of faces, objects, and letters (Ginsburg, 1978, 1980). He found in all cases that bandpass filters of 2 octaves in width are necessary for identification of items and that filters centered on 2 c/o (thus covering the range of 1 to 4 c/o) produced sufficient information for high-accuracy recognition. Ginsburg suggests that this range provides the general form of the objects. He argues from this that low spatial frequency information is most important for a majority of visual tasks and that high spatial frequencies are largely redundant.

In response to this, other researchers were quick to point out the relevance of high spatial frequencies to recognition. For instance, Fiorentini et al. (1983) showed that faces high pass-filtered above 5 c/o can be recognized better than those low-pass filtered at the same cut-off frequency. Furthermore, they show that high-passed images with a cut-off frequency of 8 c/o are recognized as well as complementary low-passed images, despite the latter having greater energy content. They argue from this that high frequency information is not redundant to face recognition, but is rather sufficient on its own to produce correct identification.

Hayes et al. (1986) provided interesting data that can paradoxically be seen as supporting the importance of both low and high spatial frequencies in face recognition. They had subjects match 1.5 octave-wide band-passed images in photographic positive and negative. Identification of negatives was inferior to that of positives only when the images contained low spatial frequencies, suggesting that it is the disruption of these which causes the photographic negative effect. This in turn suggests that these frequencies are highly important to face recognition. However, performance was best in both conditions when the images' spatial bands centered around 20 to 25 c/o, suggesting that these higher frequencies provided the most information for the task. In this sense, Hayes et al.'s (1986) findings suggest that a wide range of spatial frequencies are involved in face recognition. These results are compatible with the notion that different mechanisms are acting at different spatial scales and suggest that higher-order contour seeking mechanisms may be active at higher spatial frequencies. Although the researchers make no explicit statements about the purpose of low-frequency mechanisms, the implication is that they are involved in shape-from-shading.

Sergent (1986) attempted to reconcile the apparently contradictory findings of the Ginsburg (1978; 1980) and Fiorentini et al. (1983) studies (which suggested that low and high frequencies, respectively, were of paramount import in visual recognition) by noting that whereas the former used a matching task, the latter used a recognition task. In a matching task, subjects have continuous access to sample images of the stimulus faces whereas in a recognition task they are asked to first learn faces and then identify them later. Sergent (1986) argues that recognition relies more on the internal features and suggests that these are more dependent on higher spatial frequencies, whereas matching relies on the shape of the face and is thus low-frequency dependent. Sergent's (1986) position is that low frequencies, because they are processed faster, will tend to be predominant in early visual processing tasks such as matching or entry-level categorization, and that higher frequencies will predominate in later or higher visual processes such as identification. Unfortunately, this position fails to explain Hayes et al.'s (1986) findings that higher spatial frequencies were most useful in a matching task.

Contrary to the work discussed so far in this section, Parker et al. (1996) found evidence that high frequencies were equally relevant to both face and object recognition tasks. In a series of experiments they investigated the role of fine and coarse spatial information with reference to two broad theoretical positions concerning the role of these bands in higher-level vision. One position, exemplified by the work of Ginsburg (1978) and later Sergent (1986), holds that low-frequency spatial information has a special role in higher-level processes such as recognition and matching. In this theory, the processing speed advantage of low-frequency information is thought to reflect its role in guiding the analysis of higher-frequency information in a coarse-to-fine feed-forward system. It is assumed that representations in memory follow a similar model and are accessed or compared to input in a coarse-to-fine manner as well. But as Parker et al. (1996) point out, this assumption does not necessarily follow from the speed advantage of processing. It may be that while visual input is processed in a coarse-to-fine manner, stored information is not. This position, exemplified by the work Marr (1982), Biederman (1987) and Lowe (1987), holds that object representations are in a form which does not retain spatial frequency information but is rather feature-based. For these researchers the low spatial frequencies hold no special place in the recognition task. Instead, high spatial frequency edge information may be of higher importance because it defines important edges and boundaries.

Parker et al. (1996) argue that if coarse information guides the analysis of finer information, then cueing full-bandwidth images with low-pass images should have greater effects on performance than cueing with high-pass images. If, on the other hand, object representations are edge-based, then high-pass images should have greatest effect. They explore these possibilities in an experiment that combined a sequential matching paradigm with a cueing paradigm. Subjects were shown a full-bandwidth target face for 500 ms, followed by a 1 second blank and then a 500 ms comparison stimulus. In the test conditions, the latter was temporally divided into a 100 ms priming image followed by a 400 ms probe image. The probe always consisted of a full bandwidth image. The prime could consist of either low-pass or high-pass images of either the same face as the probe ("relevant" condition) or a different face ("irrelevant" condition). There were thus 4 experimental conditions. A control condition was also run in which the comparison stimulus consisted simply of a 500 ms full-bandwidth presentation.

In their first experiment using this paradigm, Parker et al. (1996) examined face recognition. They found that RT was higher for high-pass irrelevant primes than low-pass irrelevant primes, suggesting that high-pass information has more of an effect on recognition. There was no difference between high pass and low-pass for the relevant prime conditions, however. In their second experiment they ran the same paradigm using cross-category object images (e.g., a funnel, a cup, etc.). They found that relevant high-pass primes produced better recognition accuracy than relevant low-pass primes. RT data support this, showing that high-pass irrelevant primes were more disruptive than low-pass ones. Parker et al. (1996) conclude from these findings that there is no evidence for the pre-potency of coarse-scale information. That is, the fact that this information is processed first does not lead to it being more important in object or face recognition. Rather, their experiments lend support to the idea that high-frequency information is more relevant, in that the presence of a high-passed image that does not match a subsequent full-band image is more disruptive than a similar low-passed image.

The studies discussed thus far all used spatially filtered images as stimuli and matching or recognition type paradigms, but a few studies have also examined the question of spatial frequency efficacy through masking paradigms. Tieger and Ganz (1979) combined faces with two-dimensional sine-wave gratings ranging in frequency from 0.54 to 3.9 c/d. Figures were not given in terms of cycles per face width, but the display subtended 10 degrees and faces seemed to fill about 80% of the width of this based on the example stimuli given, so the mask frequencies ranged from roughly 4.5 to 31 c/o. Tieger and Ganz

(1979) report that a mask of 2.2 c/d, corresponding to roughly 18 c/o, had the greatest detrimental effect on face recognition.

Moscovitch and Radzins (1987, see also further analyses in Moscovitch, 1988 based on recommendations in Bruyer, 1988) examined face recognition in a backward masking paradigm using noise, pattern and spatial frequency masks. The latter were random dot patterns that varied in their grouping distances from .5 to 24 c/d. They found no effect of varying this factor, which seems to conflict with the earlier masking study (Tieger & Ganz, 1979), although the differences in methodology make it difficult to compare the two. Also, the relevance of spacing changes in random dot arrangements to spatial frequency as it is usually defined is questionable.

Keenan, Witman and Pepe (1989, 1990) examined masking by square wave gratings with fundamental frequencies similar to those used by Moscovitch and Radzins (i.e., 0.5, 12 and 24 c/d, roughly 1.6, 38.4 and 76.8 c/o). Contrary to Moskovitch et al. (1987), they found a significant effect of the spatial frequency of the mask, with the highest frequency being less disruptive than the other two. They attributed the difference in results to a difference in the intensity of the masks used. Moscovitch and Radzins (1987) used a high intensity mask, not matched for contrast with the face stimuli, and Keenan et al. (1989) argue that the effects of mask spatial frequency may have been obscured by intensity-based masking effects.

The studies discussed so far have concerned themselves mainly with the relative import of low vs. high spatial frequencies in face and object recognition. In contrast to these, a series of studies this decade (Bachmann, 1991; Costen, Parker, & Craw, 1994, 1996; Gold et al., 1999; Nasanen, 1999) has suggested that a band of middle frequencies between 8 and 16 cycles per face is of vital importance to face recognition, and that information outside this band provides little to the process.

Bachmann (1991) tested the recognition of tachistoscopically presented faces that were coarse-quantized. He had subjects learn six faces in normal full-bandwidth images and then tested them on pixelated ones. He tested a range of quantization levels having fundamental frequencies from 7.5 c/o to 37 c/o. His results showed a sharp increase in accuracy between 7.5 c/o (where accuracy was 45%) and 9 c/o (where accuracy was 80%). Images with higher numbers of pixels did not produce significantly better performance.

Costen, et al. (1994) attempted to replicate Bachmann's (1991) results and expanded the methodology to include low-passed and Gaussian-blurred faces. In their first experiment, they had subjects learn six faces in full bandwidth normal images and then tested them with images whose high spatial frequencies had been removed in one of three ways: Coarse-quantization, gaussian blurring, or low-pass filtering in the Fourier domain. For each manipulation, three cutoff frequencies were tested: 5.5, 10.5, and 21 c/o. Both accuracy and response time were measured. For accuracy, they found a similar pattern of results for all three manipulation types: There was a small but reliable increase in performance between the 5.5 c/o and 10.5 c/o conditions, but no difference between 10.5 c/o and 21 c/o. Coarse quantization produced significantly lower performance over all, while blur and filtering manipulations did not differ from one another. Response time showed overall similar patterns, with RT higher at 5.5 c/o than 10.5 c/o. For the quantized stimuli performance was slower overall and the 21 c/o level also produced faster performance than 10.5 c/o.

Costen et al. (1994) interpret their results as compatible with those of Bachmann (1991), suggesting that information below 8 c/o is not useful for face recognition. However, they acknowledge that their interpretations are threatened by a ceiling effect. All accuracies in this experiment were above 85% and those at 10.5 c/o and 21 c/o were all above 96%. Reactions times show a pattern suggestive of a concomitant flattening out, except in the case of the coarse-quantized images, which, producing slower overall reactions, leave room for improvement between the latter two conditions.

Because of the possible ceiling effect, Costen et al. (1994), performed a second experiment using a more difficult discrimination task in which the face stimuli used were selected to be more homogeneous than in their Experiment 1. The second experiment also used slightly different cut-offs (22.5, 11.5, 6 and 4.5 c/o) than the first, and a "jumbling" manipulation was introduced in place of gaussian blurring (which is equivalent to a form of Fourier low-pass filtering in any case). Jumbled faces had their Fourier components slightly randomized in terms of orientation, phase and amplitude. This retained the energy distribution of the face elements above the cut-off, but destroyed their two-dimensional structure. Accuracy data in this experiment again revealed a discontinuity around 8 c/o. In general, performance was similar for the two lower cutoffs, which differed from the two higher cutoff. The latter did not differ significantly between themselves. Accuracy data in this case remained below 80% in all cases, assuaging concerns about a ceiling effect. Overall, the findings support the idea that there is a discontinuity in recognition performance as one lowers the low-pass cutoff of a face image below 8 c/o.

In a second study Costen et al. (1996) went on to determine if there was a discontinuity in performance as one high-pass filtered face images. They had subjects recognize faces that were either high-pass filtered, low-pass filtered or pixelated. The cut-offs used were 4.5, 6, 11.5, and 22.5 c/o in all cases. Subjects learned to identify six faces in full bandwidth format and then were asked to identify these in the three different filtered modes. Results for the low pass and pixelization conditions essentially replicated those of the earlier study. Results for the high pass conditions showed a discontinuity in accuracy between the 22.5 c/o cut-off and the others, suggesting a critical threshold somewhere between 11.5 and 22.5 c/o. Based on this and other results (Bachmann, 1991; Costen et al., 1994; Fiorentini et al., 1983) they propose that a band between 8 and 16 cycles per face is critical to face recognition and that information outside this band is of little importance to the process.

Nasanen (1999) supported the idea of a critical band in a study using various manipulations of spatial frequency information in face images. He compared the performance of human observers to that of an ideal observer. Briefly, an ideal observer is a computer algorithm designed to make perceptual decisions based on optimal use of the information in the images it is given. The ratio of human performance to machine performance is termed "efficiency" and is a measure of how well humans use the information in the image in making their perceptual decisions. In Experiment 1, Nasanen (1999) showed that masking by narrow-band noise produced the greatest effects (on human observers relative to the ideal observer) when it is centered around 11 c/o. Sensitivity calculations further suggest that most of the contrast energy is gathered from a band 2 octaves wide centered around this point. Experiment 2 was a replication of Experiment 1 in which learned images were not provided to the human subjects. This was done to show that the effect is truly one of face identification and not just image matching. Experiment 3 examined the effect of phase-randomizing a narrow band of spatial frequencies and finds once again that the greatest effect occurred when a band of middle frequencies (around 8 c/o) was disrupted in this way. In a final experiment, Nasanen showed that human observers have lower contrast energy thresholds at middle spatial frequencies (between 6 and 14 c/o).

Gold et al. (1999) ran a similar set of experiments to those of Nasanen (1999). Again the researchers compared human and ideal observer performance, but they did so with two types of complex pattern: Letters and faces. In their first experiment, they compared the contrast variance threshold of human observers and an ideal observer in recognition of faces narrow-band filtered to 1 octave in width. They found surprisingly that human observers were completely unable to perform the task except when the center frequency of the band was 17.5 c/o (though one observer could perform above chance at 8.8 c/o). Under the same conditions, letters could be identified above chance across almost the entire range of center frequencies tested (1.1 c/o to 70 c/o). This difference is quite surprising considering that ideal observers were little affected by changes in the center frequency of the filtered images and could perform above chance in all conditions with both face and letter stimuli. Similarly, in a second experiment, Gold et al. (1999) found that with a 2-octave wide band, human observers fared much better, and were able to identify faces above chance in all but the lowest center frequency conditions. Efficiencies for letter stimuli show a peak around 6.2 c/o. Maximum performance with face stimuli is also around 6.2 c/o, but because humans could not perform the task below this center frequency it is difficult to ascertain if the maximum represents a peak of a band-pass shaped function. Overall, however, differences in efficiency across conditions were small, with only a 0.25 - 0.5 log unit change in efficiency across conditions for letters and 0.5 to 1.2 log unit change for faces. These data are compatible with the critical band hypothesis, although the small magnitude of the differences and the missing data points make it difficult to conclude this with confidence.

The literature reviewed in this section uses a wide variety of methods and shows an equally varied range of results. Particularly in the case of object recognition, there seems to be little consistency in findings, with some studies suggesting the importance of low spatial frequencies and others high spatial frequencies. In face recognition there is somewhat more agreement that middle frequencies are of higher importance, although even here there are some contradictory findings (Parker et al., 1996; Tieger & Ganz, 1979; Moscovitch & Radzins, 1987). Sergent (1986) attempted a reconciliation of the disparate results of earlier studies (Ginsburg, 1978, 1980; Fiorentini et al., 1983), invoking different spatial information requirements for different tasks. Her position was that matching tasks use low spatial frequencies whereas recognition tasks use high frequencies. Unfortunately for her argument, studies using matching paradigms have found support for both low frequencies (Hayes, et al., 1986; Hayes, 1988) and high frequencies (Parker et al., 1996). Thus, this explanation on its own cannot explain the differences in results.

One factor that has been largely ignored in literature on spatial frequency effects is that of spatial frequency overlap. In most of the studies mentioned above, subjects were asked to match full bandwidth pictures with either band passed images or high-passed/low-passed images of successively narrower bandwidth. In the former case overlap is kept constant and in the latter it varies concomitantly with the image bandwidth. Exploring the overlap factor in isolation requires testing subjects' ability to match two filtered images to one another, while manipulating the similarity in bandwidth between the pair of stimuli. Although a few studies have examined the ability of subjects to match filtered images to one another (Millward & O'Toole, 1986; Biederman & Kalocsai, 1997; Kalocsai & Biederman, 1998), none has varied the similarity of bandwidths occupied by the images. The few studies that have looked at matching pairs of filtered images have done so with the goal of understanding the nature of the stored representation of visual objects. The following section reviews this topic.

Models of Spatial Frequency Content in Face and Object Representations.

Much effort has been devoted to determining the cognitive and perceptual nature of face recognition (Burton, 1994; Bruce & Young, 1986; Young & Bruce, 1991; Burton, Bruce, & Johnston, 1990; Bruce, Burton & Craw, 1992). The aspect of this that is of the most relevance here is the nature of face and object representations and how they use (or do not use) spatial frequency information. Only a few studies have examined this question explicitly. These are described hereunder.

Millward and O'Toole (1986) were the first to examine the question of matching images with limited spatial frequency overlap. They found that faces low-pass filtered at 11 c/o could be matched to spatially complementary images with fairly good accuracy (63% with one learning run, up to 72% with three learning runs). They also found that subjects were better able to match similarly filtered images to one another than to full-bandwidth images. Millward and O'Toole (1986) explain this pattern of findings by postulating a form of "common information" contained in all versions of the images. That is, spatially complimentary images can be matched to each other because similar spatial-frequency-free features are extracted from them. This idea is based on Marr's (1982) influential feature-based model of spatial vision, especially relating to his "spatial coincidence assumption", which holds that spatial frequency information is correlated across scale in natural scenes.

This assumption been supported by empirical and formal investigations of natural images (Watt, 1988, 1991; Wandell, 1995).

Because information is correlated across scale, a feature-based model of vision such as Marr's should extract similar information from low-passed and high-passed images. Using a least-squared error model on their data, Millward and O'Toole (1986) find support for their explanation, as the analysis reveals that the primary source of information being used by subjects to make matches between the high- and low-passed images is indeed a form of common information. This information is found to be much more important than information contained solely in either image. Millward and O'Toole (1986) hypothesize that this common information is equivalent to Marr's (1982) "primal sketch", which is a representation holding only image features and that is free of spatial scale.

In an experiment somewhat similar to Millward and O'Toole's (1986) in concept if not implementation, Biederman and Kalocsai (1997) examined priming and naming of images that were complimentary in the Fourier domain. However, rather than simply high-passing or low-passing their images, they implemented a more unusual form of spatial complimentarity. This involved dividing the Fourier domain into an 8 x 8 array of frequency by orientation cells, similar in appearance to a chess board. One image in a complimentary pair had every odd diagonal of cells -- analogous to the black squares on a chess board -- filtered out of them, while the other had every even diagonal of cells -- analogous to the white squares on a chess board -- removed. The purpose of this research was to examine how faces and objects are stored in memory. The researchers hypothesized that faces are encoded as arrays of activation values derived from low-level oriented wavelet filters. The activation patterns are stored in the form of a metric space, a format that retains all spatial frequency information from the input image. In contrast, objects are stored as Geon Structural Descriptions (GSDs), a format that does not retain spatial frequency information. GSDs are qualitative descriptions of objects derived from non-accidental features in an image. This form of representation consists of a number of volumetric primitives (e.g., cones, sphere, bricks) and statements about their relations to one another, their aspect ratios, and a few other properties (Biederman, 1987).

To support this model of different representational formats, Biederman and colleagues (Biederman & Kalocsai, 1998; Fiser, Biederman, & Cooper, 1997; Kalocsai & Biederman, 1997) cite evidence suggesting that objects are not represented in a format that retains raw spatial frequency information. That is, they argue that objects cannot be represented in a format that retains the similarity space of the activation values of spatial filters. One example of the evidence they provide is seen in Biederman's (1987) research on contour deletion. These experiments showed that when half the contour in a line drawing is eliminated so that the GSD cannot be extracted from an image -- that is, if corners and intersections are eliminated -- recognition becomes impossible. However, the same amount of contour deletion, when it does allow GSD extraction -- that is, when sections between corners and intersections are eliminated -- allows easy recognition. In a related experiment, Cooper and Biederman (1991) show that a line drawing in which every odd line and vertex is eliminated primes a complementary image, in which every even line and vertex is eliminated, as well as it primes itself. From this they argue that the visual system is treating these two images as identical even though the pattern of activation produced by the two images would be very different.

Based on the above, Biederman and colleagues (Biederman & Kalocsai, 1997; Kalocsai & Biederman, 1998) argue that a GSD-based model of representation is best for describing object recognition. However, they acknowledge that such a model is not appropriate for face recognition, which shows strong effects of rotation, contrast reversal and so on, effects that are not compatible with a GSD representation. They therefore postulate that faces and objects are stored in different types of representations.

Biederman and Kalocsai (1997) tested this theory with a number of empirical investigations. In their first experiment, they had subjects name common objects in two blocks of trials. In the first block subjects were simply asked to name objects that were filtered as described above. In the second block, they were asked to name either the complementary image of the same object, the identical image of the same object, or an image of a different object with the same name (e.g., a different shaped chair). They found that different exemplars with the same name were named more slowly on the second block than were complementary images of the same exemplar, and that there was no difference between the complementary and identically filtered images. Thus, complementary images of objects are treated virtually the same as the identical images of those objects. Only an actual change in the GSD representation causes an effect. This insensitivity to spatial frequency contents suggests a similar effect to that seen in Cooper and Biederman (1991), but instead of outline deletion spatial-element deletion is at work.

A second experiment conducted by Biederman and Kalocsai (1997) examined name verification of famous faces. Subjects were first shown a name of a famous person, followed by an image of a famous person. If the face and name belonged to the same individual, they answered "same", otherwise they answered "different". Analogously to the first experiment, the second image shown could be the identical image, the complement of the same image, or another picture of the same person (with a different pose, expression, etc.). In this case, the results showed that complementary images were named significantly more slowly than identical images, and that the difference between complementary images and different exemplars was non-significant. Thus, contrary to the results with chairs, complementary and identical images were treated quite differently. In fact, complementary images were responded to in virtually the same way as images of completely different people. This suggests that face perception is highly sensitive to the original filter values that are input, and that face representations must therefore retain information about spatial frequency.

In summary, Biederman and Kalocsai's (1997) results showed that performance in matching identical and complementary chair images was very similar, whereas subjects did significantly worse matching complementary faces than identical ones. This supports their hypothesis that object representations are spatial-frequency free whereas face representations are not. The matching of chair images shows an invulnerability to changes in filter input values, while faces show sensitivity to this factor.

Conclusion and Rationale

In the preceding literature review, I examined four issues relating to spatial frequency effects in face and object recognition. Two antecedents of modern interest in these questions were reviewed, specifically research on coarse-quantization of images and research on matching between display formats. The former was seminal in producing interest in spatial frequency factors with regards to higher-order recognition processes. This research suggested that low spatial frequencies were sufficient for recognition, and that these could be masked by structured high-frequency noise. Research on recognition across display formats was also important in generating interest in this area. The finding that line drawings are sufficient for efficient object recognition but not for face recognition was important in that it presented the possibility that the two processes relied on different bands of spatial frequencies. Some studies suggested a vital role for low frequencies in face recognition, but it is difficult to come to precise conclusions regarding this, due to the nature of the stimulus manipulations used.

Research directly concerning itself with spatial frequency effects in face and object recognition has primarily been a recent phenomenon. The results in this area have been conflicting, especially with regards to object recognition where different studies have argued for the sufficiency or necessity of widely different ranges of spatial frequencies. Similar research on face recognition has been somewhat more in agreement, with a number of studies pointing to a band of middle spatial frequencies as being most important, although even here there has been inconsistency. Finally, there has recently been a suggestion that faces and objects are stored in different representational forms, with the latter being stored in a spatial-frequency-free format that gives objects relative invulnerability to variations in spatial frequency content. This last area presents an interesting new direction for research into spatial frequency, tying as it does the perceptual to the cognitive.

Overall, the literature on spatial frequency and visual recognition appears to be somewhat confused. Past efforts to bring order to the findings of the different studies based on methodological differences have failed (Sergent, 1986) and a consideration of other factors that might contribute to the inconsistency of the findings seems necessary. Two factors that have received little attention and which might provide some help in disentangling the literature are discussed hereunder. One is stimulus calibration and the other is spatial frequency overlap.

Calibration here refers to a procedure, generally applied to computer monitors, that compensates for a non-linear relationship between the numerical value assigned to a pixel in a stimulus image and the luminance output by the apparatus presenting the stimuli. The most commonly used apparatus is a CRT. These generally show a curvilinear relationship between these two factors, referred to as a gamma function. To compensate for the gamma function, modern research typically employs either a special monitor that has been designed for linear output or, more commonly, a system wherein the available luminance output levels are measured and sub-sampled to create a colour look-up table (LUT) that is linear in nature. The LUT is a mapping between the available output values and a linear set of luminance values. Where the monitor's luminance steps are too small (normally at the low end of the range) the mapping skips over several values until the appropriate amount of increase in luminance is achieved. Where the monitor's luminance steps are too large (normally at the high end of the range), several LUT entries will contain the same numerical value so as to compress the rise into linearity.

The reason this is necessary is that the mathematical procedures applied to images to produce spatially filtered versions work on the numerical values in the image array. In a typical experiment, the images will be in 8-bit greyscale format and are thus input to filtering algorithms as two-dimensional arrays of values from 0 to 255 (equivalently, they may be normalized to a 0 to 1 scale). The output of the filtering algorithm will typically be in the same format. However, if the range of input and output values do not represent equal luminance steps (e.g., if the difference in luminance between values 0 and 1 is not the same as that between value 254 and 255) then the output image will contain spurious spatial frequency elements in the lower range, elements that are not from the original and which could hamper recognition.

Unfortunately, early researchers seem to have been either unaware or unconcerned with this problem, presenting stimuli on non-calibrated monitors or using photographs of filtered images generated on computer monitors as stimuli. Photographs are known to show a non-linear relationship between the input luminance and the reflectance (or transmittance in the case of slides) of the image that is generated. This may present difficulties in comparing the findings of past studies with those of modern ones, although the extent of these difficulties depends on the magnitude of the effects that calibration vs. non-calibration has.

Oddly, although there have been a number of papers published on the topic of calibration (Metha, et al., 1993; Olds, et al., 1999; Pelli & Zhang, 1991; Tyler & McBride, 1997), none has explored the actual effects on human performance of using calibrated vs. non-calibrated stimuli in experiments involving complex stimuli. If the effect is large, then this causes problems with comparing past studies to present ones. If it is small, then one can more confidently contrast older studies with new ones.

Although the issue of stimulus calibration is of important practical interest, a more deeply theoretical concern is found in the concept of spatial frequency overlap. As noted before, this factor has not been much examined. This despite the fact that it is likely contributes a significant portion of the effects seen in experiments examining display format as well as those examining the effectiveness of different spatial bands in recognition. For instance, there has been little consideration of the fact that as one asks subjects to match a full-bandwidth image to filtered images with ever-narrowing bandwidths, there is a loss of common bandwidth for the visual system to work with. In this sort of task, used by a number of studies (e.g., Costen et al., 1994; 1996), the loss in overlap is concomitant with the loss of information in the filtered image and it is difficult to say what proportion of effects can be attributed to one factor or the other. Separating out the two factors requires testing subjects' ability to match two filtered images with one another while varying their range of shared spatial frequencies.

Casual observation reveals that it is more difficult to match a high-pass image to a low-passed one than it is to match either one to a full-bandwidth original. But as the cut-offs of the two filtered images approach one another, they become easier to match. The reason for this is presumably that the range of the spatial spectrum they share increases as one does so. Likewise, two band-passed images become easier to match as their cutoffs become more similar. The lack of data on matching pairs of filtered images is unfortunate, as such studies may be highly informative. For instance, if middle frequencies are more useful for face recognition, then matching two images band-pass filtered so that they contain the same range of frequencies within this range should be relatively easy (e.g., matching two images filtered to have frequencies between 8 and 16 c/o). Conversely, matching two face images filtered so that they contain the same width of frequencies outside the critical band should be more difficult (e.g., matching two images filtered to have frequencies between 4 and 8 c/o). This possibility has not been examined. Studies have only examined the ability of subjects to match filtered images with unfiltered ones.

At this point, it is important to be quite clear about what is meant by the term "spatial frequency overlap". Figure 1 illustrates this concept, showing the gain profiles of filters through which pairs of stimulus images are passed. As Figure 1a shows, when two images are filtered — in this case when one is high-passed and the other low-passed — they will vary in the range of spatial frequencies they share. If the low-pass cutoff is relatively high and/or the high-pass cutoff is relatively low, the range of shared bandwidth will be high. As the high cutoff rises and/or the low cutoff falls, the shared portion of the

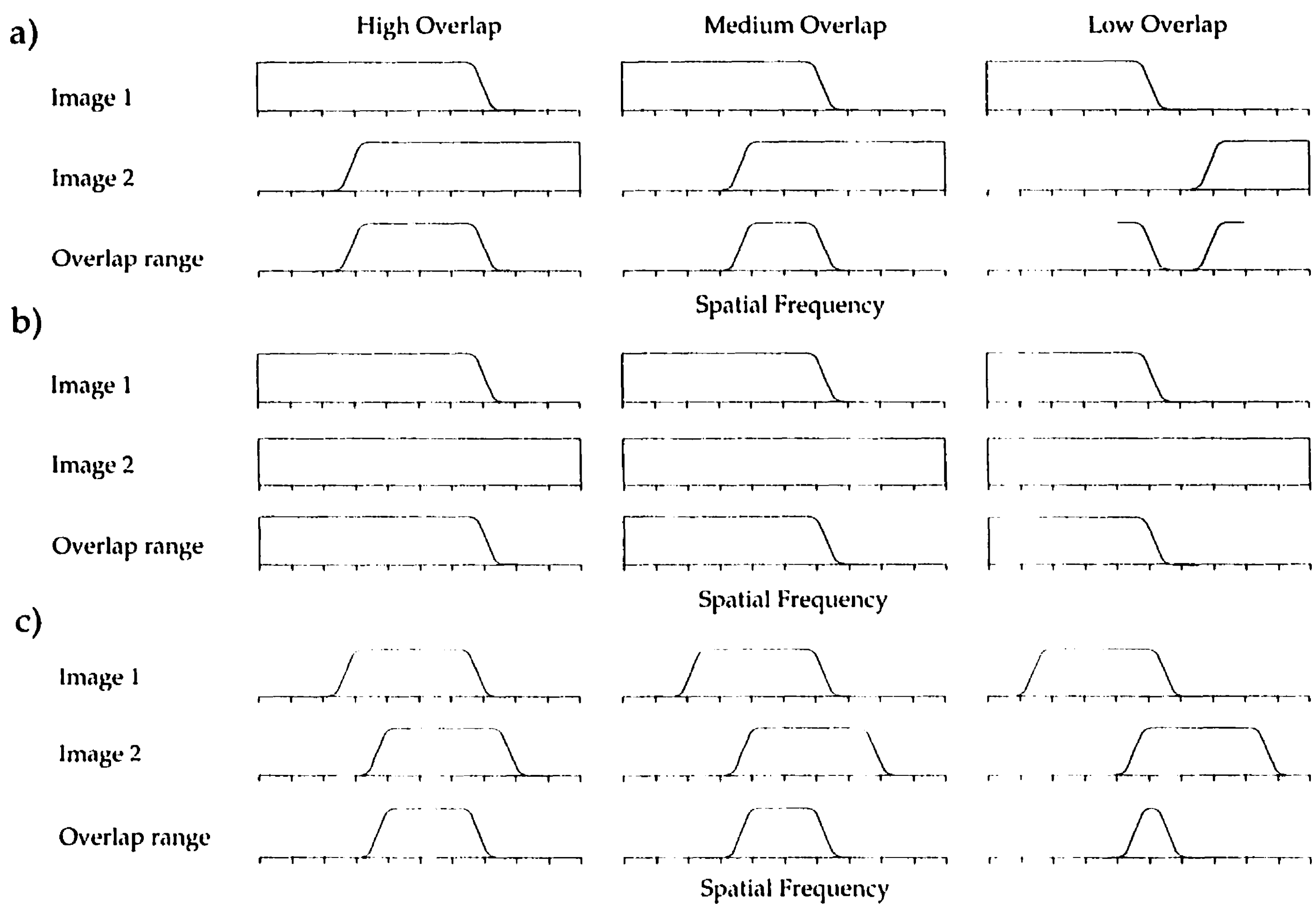

Figure 1. An illustration of the concept of spatial frequency overlap. a) Variations in overlap as a pair of images are low-passed and high-passed at successively more extreme cut-off frequencies. b) Overlap varies as a full-bandwidth image is compared to a filtered image with a successively more extreme cut-off. c) Two band-passed images share less common spatial frequency bandwidth as their center frequencies are seperated.

spectrum becomes smaller and may become zero. As the cutoffs continue to separate, the range of shared spatial frequencies may be said to be negative. Throughout this dissertation, spatial frequency overlap is given in octaves of spatial frequency. When two images share a portion of the spectrum, overlap is positive. When two images are separated, overlap is negative. Negative levels of overlap might be expressed in positive terms as a number of octaves of "separation" or "gap" between the images, but for consistency only the term "overlap" will be used throughout.

Changes in spatial frequency overlap are also seen when one of a pair of images is unfiltered and the other image is successively filtered to narrower spatial ranges. This is illustrated in Figure 1b. Here, as a low-passed image is restricted to an ever smaller range of spatial frequencies, the range of common spatial frequencies it shares with a full-bandwidth image shrinks concomitantly (i.e., the shared range is identical to the range of the filtered image. Finally, as illustrated in Figure 1c, if two images are band-passed at ever closer center frequencies, they come to share a greater portion of the spatial spectrum while the total range of the images remains constant.

To summarize the above, spatial frequency overlap can be defined as the range of the spatial scale simultaneously occupied by two images in a learn/test pair or in a pair of images for matching. This factor is independent of the absolute position on the spatial scale, as a given degree of overlap can be produced at any point on the spectrum and with any bandwidth of image (within the limits of the fundamental and Nyquist frequencies of the image, of course).

In some sense what is being presented in Figure 1 is a subset of a four-factor problem, where a full illustration would show a range of conditions manipulating the low- and high-pass cutoffs of each image in a pair independently. Some parts of this four-dimensional parameter space are obviously not of interest, for instance where one or both images contain no spatial information because their low cutoff is lower than their high cutoff. Other portions of it are of limited interest as well, as when the two images are too far distant on the spatial scale to be matched even with optimal use of the data present. Exploring this space in terms of the factors of spatial frequency overlap, image bandwidth, and center frequency is fruitful because it necessarily keeps one within the informative and interesting parts of the total parameter space.

The value of exploring recognition in these terms can be seen by examining past studies that have supported the notion of a critical band of spatial frequencies for face recognition (Costen et al., 1994; 1996). These studies have subjects recognize or match images that have been low- or high-pass filtered to learned images that are full-bandwidth. The cutoffs are varied and a sudden drop is seen around 8 c/o for low-pass test images and 16 c/o for high-pass test images. This situation is analogous to that in Figure 1b, where the entire bandwidth of the tested image is contained in the learned image but not vice-versa. Although one interpretation of these results is to say that the more informative frequencies are between 8 and 16 c/o, several other possibilities remain. For one, it is possible that the narrowing of the spatial band in the filtered test images is contributing to the sudden drop off. Thus, a band-passed image with a three-octave width might produce similar performance to images with frequencies of 1 to 8 c/o, no matter what the center frequency of the band is.

Another alternate hypothesis is that interference is occurring because of frequencies in the unfiltered image that do not match those in the filtered image. It may be that making two images more similar in their spatial range will result in improved performance regardless of position on the spatial scale. That is, spatial frequency overlap may play a role in these findings. Consider, for the sake of illustration, a simplistic template-matching model of visual recognition wherein the visual system is performing a feature-by-feature match between an array of early wavelet-detector activation levels in an input layer and a similarly-arranged stored image. If the stored image is full-bandwidth and the input image is filtered, a number of factors might progressively impede recognition as the input image's bandwidth is narrowed. The most obvious is that there will be fewer "hits", that is fewer elements in the input image that match elements in the stored image. It is also true that there will be more "misses" as there will be a progressively larger number of elements left in the stored image that are not in the input image. But if the stored image is also filtered, then the number of misses is reduced while the number of hits remains the same. To put it succinctly, as the stored image is filtered to occupy a similar spatial frequency range to the input image, the correlation between the two images increases. Effectively then, filtering out the non-overlapping frequencies reduces noise in the matching process.

Of course, the noise in this case is not noise in the traditional sense, as it is not random. Instead, it correlates with the signal (i.e., the elements in the stored image that match the input image). This bears an interesting resemblance to the problem of coarse-quantized images, where a structured form of "noise" at higher spatial frequencies is thought to mask lower-frequency face structure. It was argued earlier that the pixelization effect is not really a signal vs. noise problem but rather a problem of two competing signals. In both cases, because the noise is structured, it presumably competes with the underlying image. In the case of coarse-quantized images, the competing signals do not correlate and are thus highly disruptive to each other. The grid pattern, being the more structured element, wins out in terms of perceptual saliency. Only eliminating this competing signal by low-passing the image (Harmon, 1971; Harmon & Julesz, 1973) or disrupting its structure by adding energy at oblique frequencies causes the other signal to emerge (Morrone et al., 1983). In the case of matching a filtered face to a full-bandwidth face the competing signals are correlated, so the interference is lesser. Indeed, elements close in spatial frequency should aid recognition, but elements distant enough on the spatial spectrum should act as noise in the recognition process because their correlation with the elements to be matched is insufficient.

The above is not to say that some bands of frequencies are not more informative than others for a given task, indeed this seems likely to be true. But although information on these bands provides useful information concerning human recognition systems, it is not sufficient to simply know the efficacity of different bands if one wishes to predict human performance in certain recognition tasks. For instance, when the spatial frequency content of neither image is wholly contained within the other, such as when one must recognize a low-pass filtered image from a high-pass one or (more naturalistically) when two different display formats are being used, a very different situation from that examined in the critical band studies arises. In this case, knowledge about the critical band alone may not provide enough information for accurate predictions of recognition performance.

It is fairly obvious that recognizing a low-pass image from a high-pass one (or vice versa) will generally be more difficult than matching either of them to a full-bandwidth one. This shows that even if a pair of filtered images have sufficient information in common with an unfiltered one to be matched with it, they may still not have sufficient information in common with one another to be recognized as representations of the same object. This idea is empirically supported by the work of Millward and O'Toole (1986) who show that matching low-passed or high-passed face images to similarly filtered images allows better performance than matching either of these to a full-bandwidth image. The researchers also show that matching a low-pass image to a high-pass one (or vice versa) is more difficult than any other condition. These findings are important in that they support the importance of spatial frequency overlap in visual recognition.

Also illustrative of the importance of this factor is an examination of some of the contradictory findings in the literature concerning which spatial frequency bands are most useful in face recognition. For example, Costen et al. (1994, 1996) had subjects recognize images that were low-passed or high-passed at progressively more extreme cut-offs while the learning stimuli remained full-bandwidth. In this situation it is not only the cut-off, but also the range of spatial frequencies and the degree of overlap between learned and tested faces which changes across conditions. Using this method, Costen et al. (1994, 1996) find that the most useful frequencies for recognition are centered around 12 c/o. By comparison, Hayes and colleagues (Hayes et al., 1986) had subjects match band-passed images with full-bandwidth photos. This has the effect of keeping the overlap and range of frequencies in the test images constant. Using this methodology, they found that the most useful frequencies for recognition center around 20 to 25 c/o. The large difference in the findings of these two studies may be attributable to the fact that overlap varied in one while staying stable in the other.

An understanding of how variations in spatial frequency overlap affect recognition may also aid in comprehending recognition across different forms of representation, such as line drawings, low-resolution images, two-tone images and so on. Each of these preferentially presents information at different bands of frequencies. A line drawing, although it contains energy across the spectrum, primarily displays high-frequency information from the original (indeed a reasonably effective algorithm for computer-generated line drawings involves high-passing an image and then bi-quantizing it). In contrast, a low-resolution image retains low spatial frequencies. In situations where a person must recognize a face between two such representations, for instance when a police officer must match a forensic artist's sketch to a security video, the visual system is challenged by a lack of common information between the two representations.

A recent study (Burton, Wilson, Cowan & Bruce, 1999) shows an example of this. These researchers have shown that lay subjects, as well as police officers with substantial experience in forensic identification, were extremely poor in matching unfamiliar faces taken by a low-quality security video to the same faces presented in high-quality photographs. Part of the difficulty may be attributed to the discrepancy between the spatial content of the two kinds of images. The high-quality photographs contain frequency elements that are not present in the low-quality video images. Recognition should be even harder or impossible in an example where a face in a low-resolution video image must be matched with a sketch. In this case, neither image wholly contains the spatial frequency range of the other and in fact, the two images may not share any of the spatial frequency spectrum at all. Similarly, identifying a face learned in a two-tone image from a line drawing would be quite difficult because the two-tone image preserves low spatial frequencies while disrupting high ones (Hayes, 1988; Liu & Chaudhuri, 1997, 1998).

The examples and studies cited so far show that spatial frequency overlap is an important factor in image recognition, but to date there have been no studies examining it quantitatively. Although some information can be inferred from studies that have examined images complementary in the Fourier domain, no research has been done that varies the range of spatial frequency in common between images to quantitatively determine the effect of this manipulation.

The main purpose of this study was to systematically examine the effects of spatial frequency overlap on face and object perception by varying the range of spatial frequencies shared by pairs of images used in learning and matching paradigms. This will allow us to estimate how much transfer occurs when an object is learned through one set of frequency channels and tested through another. Within this context, we were interested in determining the point at which floor and ceiling performance would be achieved as well as determining the degree of improvement in performance elicited as images become more similar in spatial content.

Along similar lines, we were interested in determining the relative contribution of spatial range similarity as compared to position on the spatial spectrum. That is, assuming that different bands of frequencies are more efficacious for recognition than others, how much of a difference does this factor make relative to the spatial frequency overlap factor? Also, we wished to determine if spatial frequency overlap effects are independent of position on spatial scale or if there is an interaction (i.e., will the same change in overlap have a different effect for images at the high and low ends of the spectrum?)

A number of different methodologies have been used in the literature on this topic, sometimes making the literature difficult to interpret. Sergent (1986) has suggested that variations in spatial frequency information may have different effects for different tasks. For this reason, we were interested in examining how spatial frequency information affects performance in both recognition and matching paradigms.

Another factor that hinders interpretation of the literature on spatial frequency effects in visual recognition are differences in how stimuli are presented. Many researchers have presented their images on non-linearized monitors. In fact, of the studies cited in the above literature review, only one (Peli, 1992a) explicitly states that a linearized monitor was used, and another (Millward & O'Toole, 1986) clearly states that their stimuli were non-linear in nature. Others (e.g., Fiorentini, et al., 1983) used slides to present stimuli, making it unlikely that their stimuli had a linear luminance distribution. As Peli (1992a, 1992b) points out, non-linearities in luminance lead to the introduction of aberrant low-frequency elements in high-passed images. This threatens the interpretation of earlier studies and leaves open the question of how easily one can compare and contrast their results with those of modern studies using linearized monitors.

Although it is possible to ascertain what sorts of frequency elements will be introduced into an image by non-linearities, it is difficult to predict what sorts of effects these will have on subject performance in recognition tasks with complex images. In order to be able to compare past studies to present ones, it would be useful to know to what degree non-linearities affect performance. For this reason, several of the studies presented below have been performed twice: Once with a linearized monitor and once with a non-linearized one. Differences between the outcomes of these two sets of experiments should give us some insight into how to compare older and newer studies.

Though the primary goal of the studies presented below is to examine the effects of spatial frequency overlap on visual recognition, another objective is the examination of models of face and object representation, specifically the "two representation" model proposed by Biederman and Kalocsai (1997; Kalocsai & Biederman, 1998). They propose that face representations retain information about the spatial frequency of features whereas object representations do not. They argue that the spatial-frequency-free representation of objects gives them greater robustness to variations in image quality. If this is the case, then object recognition should be less affected by variations in spatial frequency overlap than face recognition, as the former will involve essentially the same GSD representations across a wide range of spatial filtering conditions, while the latter will suffer differences in representation as spatial frequency varies.

If the two representation model is correct, then we should see a greater robustness to lack of spatial frequency overlap for object recognition than for face recognition. However, there is an alternate hypothesis for results supporting this model. That is, it is possible that greater robustness arises from greater correlation of information across scale in the object images. Indeed, this seems likely as object images tend to have more sharp edges, and sharp edges transcend spatial scale to a greater extent than smooth shape and shading variations. To examine this alternative hypothesis, a control condition in which upside-down faces are used as stimuli is run. Upside-down faces are thought to be treated more like objects by the visual system (Haxby et al., 1999; Aguirre, et al., 1999), and so they should show a similar pattern to objects if that is the case. If, on the other hand, they show a pattern more similar to faces, then this argues that Biederman's findings are due to a difference in the stimuli.

To summarize, the following questions are examined in the studies presented in this dissertation: 1) What is the effect of varying spatial frequency overlap, and how does this compare in magnitude to the effect of varying the location on spatial scale? 2) What can the answer to question #1 tell us about recognizing images across display format? 3) What can the answers to questions #1 and #2 tell us about which bands of spatial frequencies are most useful for different tasks? 4) How does luminance calibration of filtered stimuli affect performance in such tasks and what does this tell us about comparing studies that do not use calibrated displays to those that do? 5) How are the effects of spatial frequency overlap and location on spatial scale affected by type of task (learn/test vs. matching)? 6) Do objects and faces differ in how spatial frequency overlap affects them and what can this tell us about the format of mnemonic representations for these two stimulus types?

Finally, the following gives a brief synopsis of each experiment's methods, rationale and predictions:

Experiment 1: This experiment looks at face recognition in a learn/test recognition paradigm. It examines the tolerance of the face recognition system for a lack of common bandwidth as well as quantifying the effects of increasing overlap at three different locations on the spatial scale. Subjects are asked to learn images in low-pass and recognize in high-pass, or vice versa. The cutoffs of the filters are separated to different degrees to create different levels of overlap. This experiment, as well as experiments 2 to 4, are broken up into two sub-experiments a and b, which examine 3 lower levels of overlap (-2, -1 and 0 octaves of overlap) and 3 higher levels (1, 2 and 3 octaves of overlap), respectively. The general prediction here is that performance will rise with overlap and this factor will be a more important determinant of performance than position on spatial scale.

Experiment 2: This experiment replicates Experiment 1, but uses an uncalibrated display to examine the comparability of older studies that did not use calibrated stimuli to modern studies that did. It is predicted that calibration will have greatest effects in conditions that involve high-pass images with a fairly extensive spatial bandwidth but that in general effects on performance will be small.

Experiment 3: This is a control and baseline experiment. Its methodology is similar to Experiments 1 and 2, but it looks at subjects' ability to match filtered images to other similarly filtered images. The goal here is to determine how much effect the actual spatial content of images has on recognition as compared to overlap. It is predicted that this factor will have little effect overall.

Experiment 4: This experiment looks at face recognition in a simultaneous matching paradigm. This allows us to generalize our data to other studies that used a similar paradigm and to examine any paradigm-dependent differences in how spatial frequency information affects visual recognition. It is predicted that overlap will have as strong an effect here as it does in recognition, though there will overall be higher levels of performance due to minimized memory load.

Experiment 5: This is a replication of Experiment 4 using uncalibrated stimuli. In a similar manner to Experiment 2, it allows us to determine the comparability of past studies to newer studies. As with Experiment 2, the prediction is that effects of calibration will be small.

Experiment 6: This experiment contrasts the effects of spatial frequency overlap in the recognition of faces vs. objects. In this case, bandpass images are used in a sequential matching paradigm. This allows us to test one prediction of Biederman and Kalocsai's (1997; Kalocsai & Biederman, 1998) theory that face representations retain spatial frequency information whereas object representations do not. The hypothesis is that face recognition will show a greater susceptibility to spatial frequency overlap than object images.

Experiment 7: This experiment examines an alternative hypothesis for results supporting of Biederman and Kalocsai's (1997; Kalocsai & Biederman, 1998) theory in Experiment 6. Specifically, it is possible that object images would show less susceptibility to spatial frequency overlap simply because the information in them is more strongly correlated across scale. To test this, the same experiment was run with upside-down faces. These are thought to be treated more like objects by the visual system (Haxby et al., 1999; Aguirre et al., 1999) and as such should show a pattern that is similar to that of objects, assuming the two representation theory is correct. If they show a pattern more like upright faces, then this argues that the effect is simply due to differences in the stimuli.

CHAPTER 2: EXPERIMENTS

This section describes seven experiments examining the effects of varying spatial frequency overlap on face and object recognition. Experiments 1 to 5 deal solely with face recognition, whereas Experiments 6 and 7 compare performance with object and face stimuli. Each set of experiments is preceded by a section describing the stimuli used in them. The data from these experiments was previously presented in two manuscripts: Experiments 1 to 5 were reported in Liu, Collin, Rainville and Chaudhuri (2000), while Experiments 6 and 7 are reported in Collin, Liu and Chaudhuri (submitted).

Stimuli for Experiments 1 to 5

The stimuli for Experiments 1 to 5 were spatially-filtered images of faces. The original images were obtained from an internet database at the University of Essex. All pictures were from the forward view. There were 46 individual faces, 10 of which were set aside as practice stimuli. All images were converted to 256 gray-level format before filtering. Filtered images were equated for mean luminance.

The spatial filtering of the images was done with MatLab 5.2 software for MacIntosh. To create the filtered versions, the original full-bandwidth images were Fourier transformed, then convolved with smooth Butterworth filters, and finally inversely transformed back into the spatial domain. Butterworth filters were used to avoid the ringing effects seen when images are filtered with abrupt spatial frequency cutoffs. Nonetheless, the functions were steep enough to provide good spatial frequency localization. The low-pass filters were defined by

$$\frac{1}{1+(r/c)^5}$$

and the high-pass filters by

$$\frac{1}{1+(c/r)^5}$$

where r is the component radius and c the cutoff radius.

Filter cutoffs were selected to cover a wide range between the fundamental frequency of the images (1 c/o) and their folding frequency (about 68 c/o). Because of the image dimensions, the highest cutoff that did not result in a truncation of one of the high-pass filters was 26.9 c/o. This was taken as the upper cutoff value. Three center frequencies were then determined by maximizing their spacing within this range, producing values of 5.3, 8.0 and 12.0 c/o. Six overlap conditions were defined at each of the center frequencies by placing low-pass/high-pass filter pairs symmetrically about them. The overlap conditions varied by steps of one octave and ranged from -2 to 3 octaves of overlap.

Each overlap level was tested at each center frequency. This was done in consideration of the fact that the same amount of overlap might have different effects on performance depending on the point in the spectrum at which the overlapping spatial frequency bands meet. This might arise because of the changing bandwidths of the stimuli as the center frequency shifts. For instance, if the bands meet at a very low frequency, the high-frequency band contains most of the spectrum whereas the low-frequency band contains only a small number of coefficients. The opposite is true if the overlapping bands meet at high frequencies. Because these differences in spatial frequency range may well have effects that are independent of or interact with those of spatial frequency overlap per se, overlap was varied about three different center frequencies. This allowed measurement of the effect of this factor along with overlap.

Figure 2 illustrates the gain profiles of all the filters used in Experiments 1 to 5. Figure 2a shows the filters used to make stimuli for the low overlap experiments (1a to 5a), which examined overlap levels of -2 to 0 octaves. Figure 2b shows the same for the high overlap experiments (1b to 5b), which examined overlap levels of 1 to 3 octaves. As overlap levels increase, the low-pass and high-pass elements of the filter pairs approach (and pass each other in panel b) in 1 octave steps. That is, at each increase in overlap the low-pass cut-off in each filter pair moves up .5 octaves as the high-pass cut-off moves down .5 octaves. The lowest cutoff frequency thus obtained is 2.4 c/o. At the maximum overlap level, the low pass filter's cut-off is 3 octaves above the high-pass cutoff. At the minimum overlap level, the low pass filter's cutoff frequency is 2 octaves below the high-pass cutoff.

In examining Figure 2, readers may note that some cutoff frequencies are used in more than one condition, representing different levels of overlap at each. This arises because the center frequencies are shifted by .5 octave steps at the same time that the low-pass and high-pass elements of each filter pair are shifted in .5 octave steps at each overlap level. Although 32 filter profiles are shown in 2 there are in fact only 20 unique filters in the diagram: 10 high-pass and 10 low-pass. These are labeled with ordinals from 1 to 10 that give an index of the breadth of the spatial frequency range removed from images processed with them. That is, filters with low ordinals let through a broader range of frequencies than filters with high ordinals. For instance, the filter ordinals 1 to 5 in Figure 2a correspond to cutoff frequencies of 12.0, 8.0, 5.3, 3.5, and 2.4 c/o width for the low pass filters and 5.3, 8.0, 12.0, 17.9, 26.9 c/o for the high-pass filters.

The relationship of the 20 filters and 18 filtering conditions (3 center frequency x 6 overlap) is somewhat complex, but can be visualized by examining Figure 3, which shows examples of the stimuli used in the experiments. Figure 3a shows examples for the low overlap experiments (1a to 5a) while Figure 3b shows examples for the high overlap experiments (1b to 5b). Each image in Figure 3 is a face image passed through one of the filters shown in Figure 2. The images shown in Figure 3a are the result of processing by

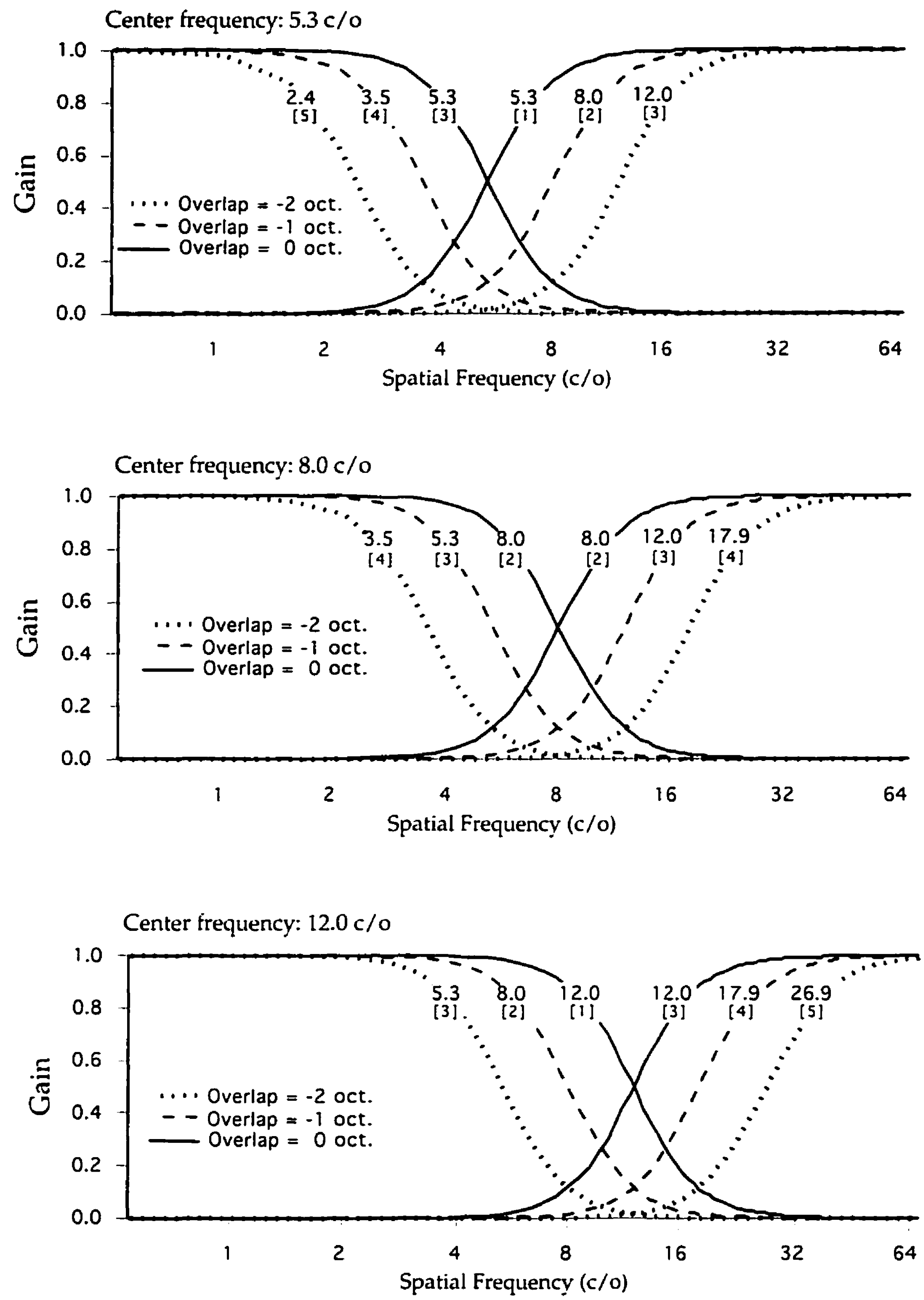

Figure 2a. Gain profiles and 50% cut-off values of filters used to generate stimuli for Experiments 1a to 5a. The bracketed numbers below the cut-off values are filter ordinals that identify each unique low-pass or high-pass filter.

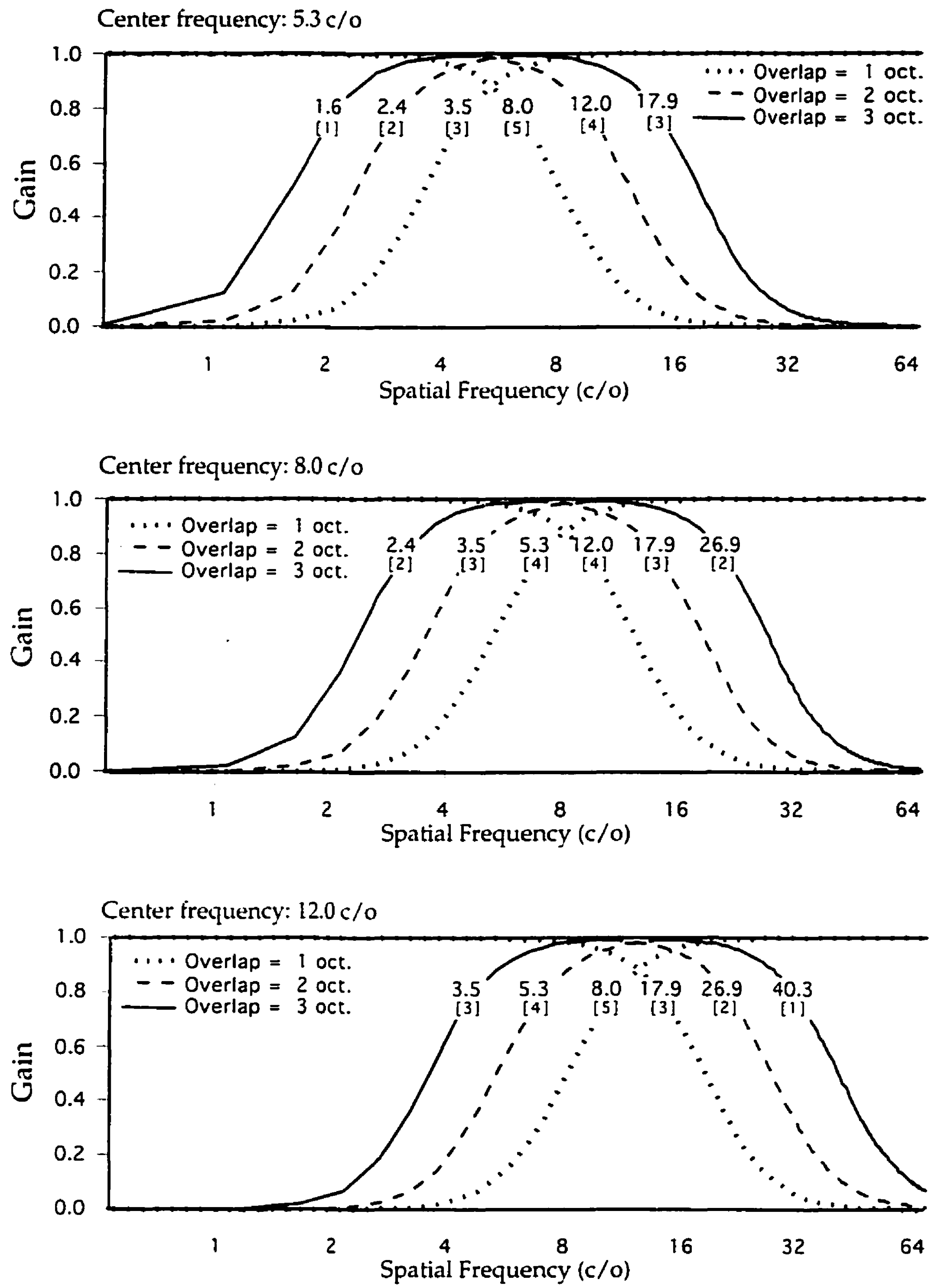

Figure 2b. Gain profiles and 50% cut-off values of filters used to generate stimuli for Experiments 1b to 5b. The bracketed numbers below the cutoff values are filter ordinals that identify each unique low-pass or high-pass filter.

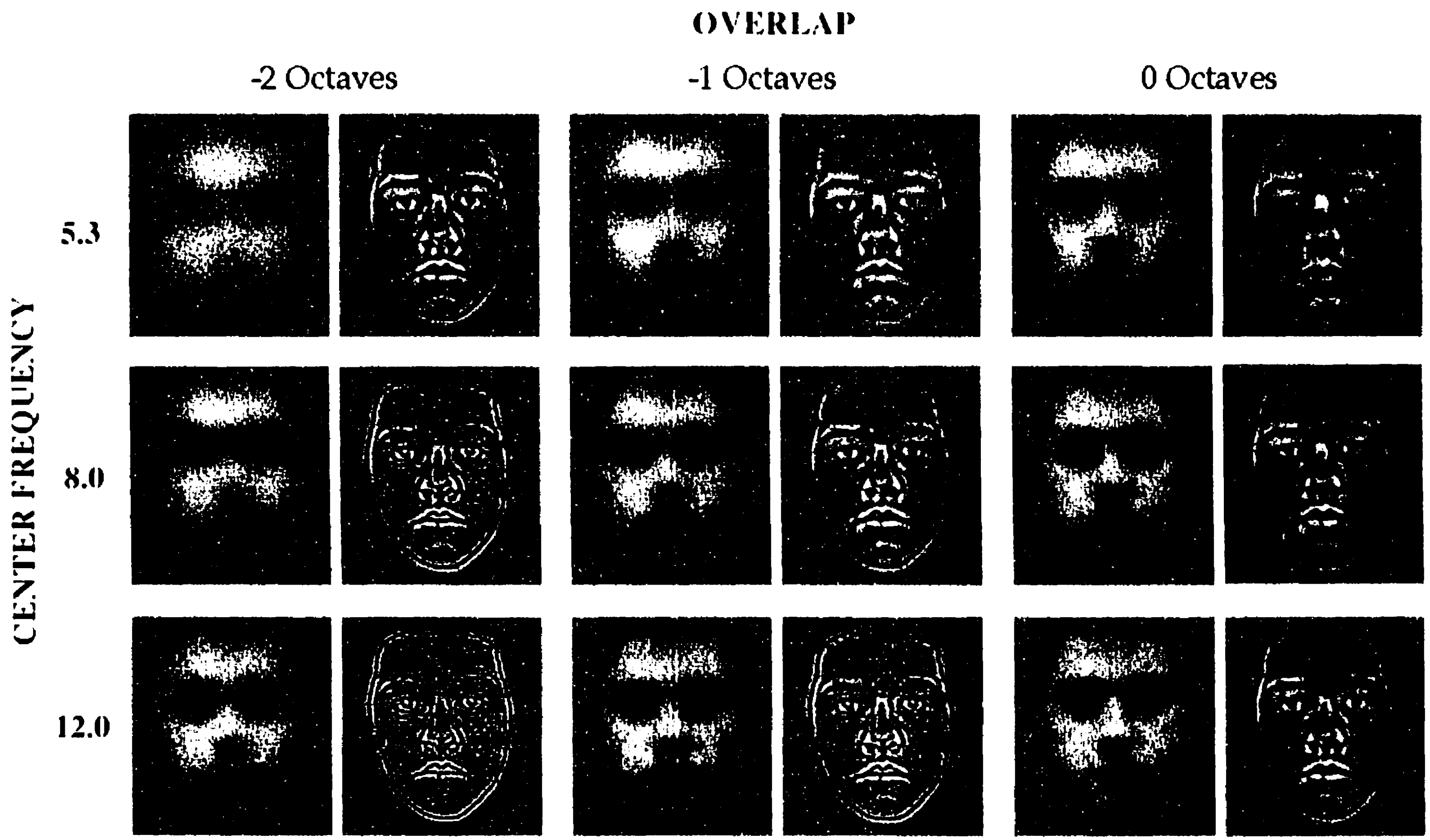

Figure 3a. Example stimuli for experiments 1a to 5a.

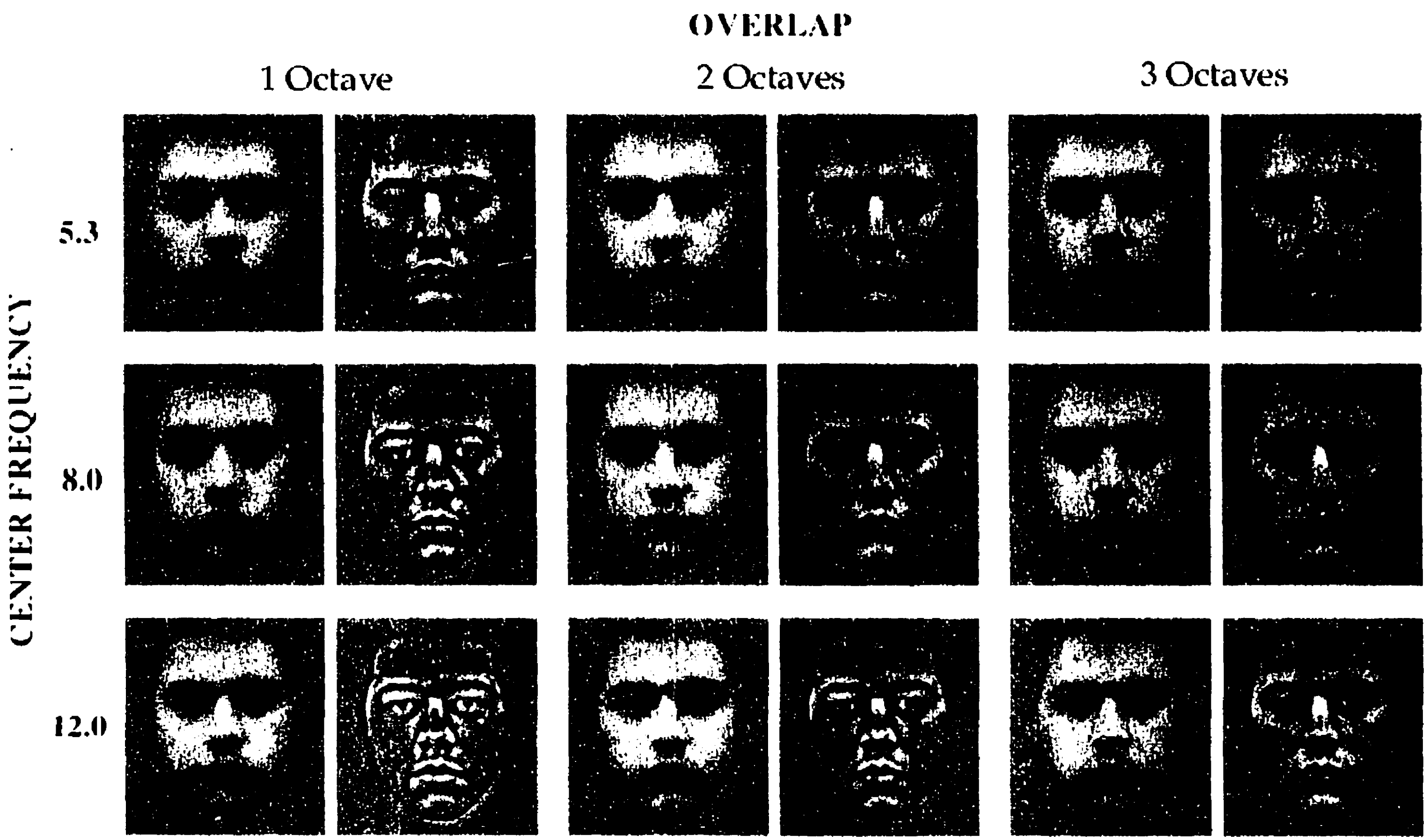

Figure 3b. Example stimuli for Experiments 1b to 5b.

filters shown in Figure 2a whereas those in Figure 3b were made by filters in 2b. The rows of Figure 3 represent the center frequencies and the columns represent the levels of overlap. In each cell of the figure, the image on the left is the low-pass version and that on the right is the high-pass version. Below the images are given the cutoff frequencies of the filters that produced them as well as the filter ordinal which corresponds to the cutoff. The total number of images used in this study was 920 (46 individual faces x 20 unique filters).

The term "opposite filters" is used throughout the report of these studies to describe filters symmetrically positioned about a given center frequency. For instance, in the low-overlap experiments the medium overlap level (-1 octaves of overlap) at center frequency 8.0 c/o involves two filters: A low-pass filter with cutoff frequency 5.3 c/o and high-pass filter with a cutoff of 12.0 c/o (see Figure 2a, middle panel, dashed lines). These are referred to as opposite filters to one another.

Experiment 1

This experiment examined the ability of subjects to recognize low-passed face images from high-passed ones, and vice versa. The distance between the filter cutoffs was varied to create different levels of spatial frequency overlap. A learn-test recognition paradigm was used.

The major goal of this experiment was to test the limits of face recognition system's ability to deal with a lack of common bandwidth between learned and tested images. Millward and O'Toole (1986) found that when the cutoffs of a low-pass and a high-pass face image are the same (i.e., the images are complimentary in the Fourier domain) subjects perform above chance in recognizing one from the other. It was therefore decided to examine a range of conditions in which the cutoffs of the filters were further apart, in addition to one which was equivalent to that in this earlier study. This was done in Experiment 1a. Also of interest was the examination of how performance continued to improve as the filter domains came to share part of the spectrum, producing positive values of spatial frequency overlap. This was done in Experiment 1b.

The general hypothesis was that performance would increase significantly as spatial frequency overlap increased, with chance performance occurring somewhere below 0 octaves of spatial frequency overlap and ceiling performance somewhere above it.

Experiment 1a

This experiment examines the effect of spatial frequency overlap on face recognition in a learn/test paradigm. The goal is to quantify the effects of varying overlap and to determine the minimum amount of overlap that allows above-chance recognition.

Method.

Participants. Twenty-one undergraduates from McGill University (7 male, 14 female) with ages ranging from 19 to 36 years (median = 20), participated. All participants had normal or corrected-to-normal vision.

Materials. Materials were as described in the general stimulus description above. For this experiment, face images were processed by the filters shown in Figure 2a. Example stimuli may be seen in Figure 3a.

Design and procedure. The design of the experiment was 3 x 3, completely within subjects. The factors were center frequency (5.3, 8.0 or 12.0 c/o) and spatial frequency overlap (-2, -1, and 0 octaves).

Subjects were tested using a Power Macintosh 7200/120 computer with a 17" monitor. The monitor was properly calibrated to correct for its gamma function. Images were surrounded by a neutral gray background that filled the screen. Instructions were given in written form on the monitor.

The initial part of the experiment was a short practice session. In the learning part of the practice session, subjects were asked to memorize 5 filtered face images presented sequentially for 4 seconds each. The faces were drawn at random from the 10 set aside for this purpose. The filter applied to each face was selected at random from the 18 possibilities (3 center frequencies x 3 overlap levels x high-pass/low-pass). Subsequently, subjects were tested to see how well they could recognize the learned faces when they were filtered with opposite filters to the ones applied to the learned faces. For each learned face, one target and one distractor were included in the testing set. Both the target and the distractor were processed by the filter opposite that through which the learned image had been passed. For example, if the learned face was *low-pass* filtered at center frequency 8.0 c/o and the -1 octave overlap condition (Figure 2a, middle panel, dashed lines), the target and distractor were *high-pass* filtered at the same levels. That is, if the learned face was low-pass filtered at 5.3 c/o, the target and distractor were high-pass filtered at 12.0 c/o.

Subjects were informed that the test images would be different from the learned images but that some of the test images were the same faces as the learned ones. Subjects were instructed to press the "Yes" key if the face had been seen during the learning phase or the "No" key otherwise. The test image remained on the screen until the subject responded.

Immediately following the practice, the actual experimental session was performed. Subjects were informed that this would follow the same procedure as the practice, but that there would be more faces to remember. During the learning part of the experimental session, 18 faces images were shown, one filtered with each of the 18 different filters (see Figure 2). Thus, 9 were low-pass images and 9 high-pass. During the testing part of the experimental session, 36 faces were presented, 18 targets and 18 distractors. As with the practice session, one distractor and one target were presented for each learned face, and these were both passed through the filter opposite that which had been applied to the learned face. The order of presentation of target faces was the same as the order in which they were learned, but with distractors randomly interspersed in the sequence.

<u>Results and Discussion.</u>

Figure 4a shows the results for this experiment. The values given are mean accuracies, with error bars representing one standard error. A two-way repeated measures

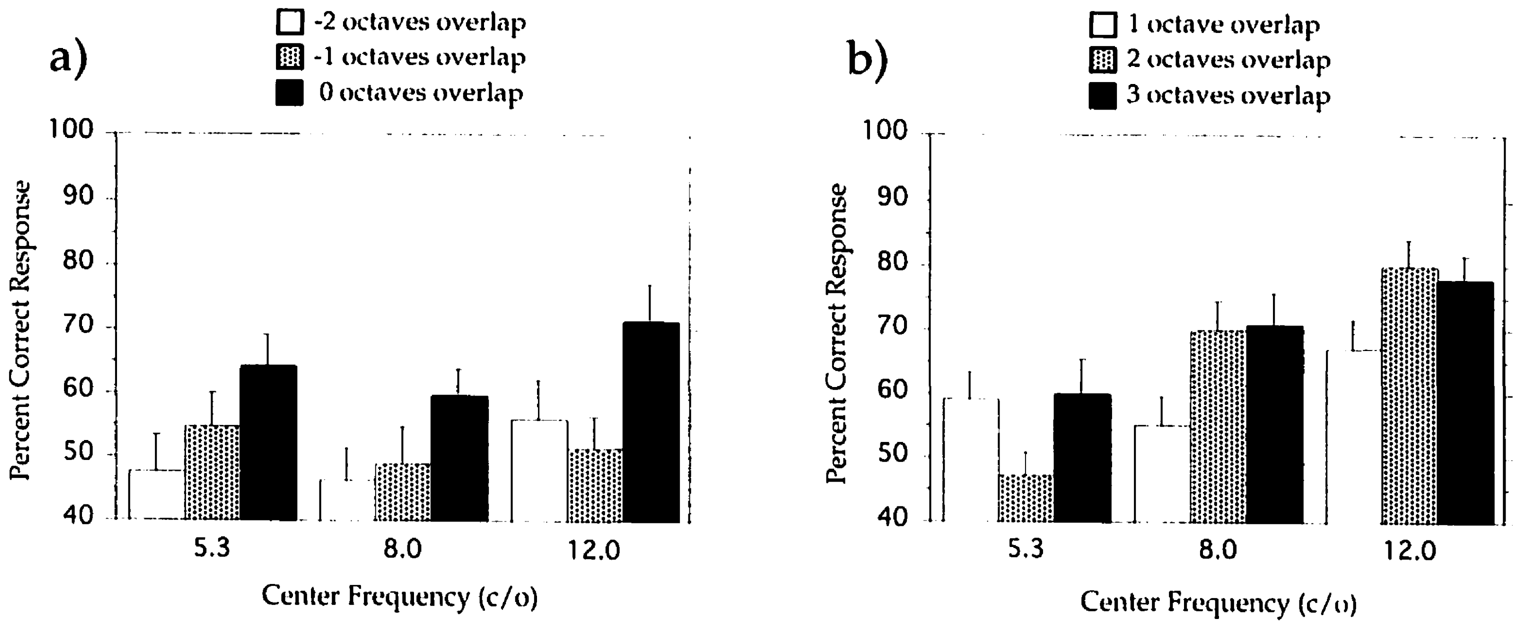

Figure 4. Mean accuracy data for Experiment 1. Error bars represent one standard error. a) Data for Experiment 1a. b) Data for Experiment 1b.

ANOVA showed a significant main effect of overlap level, $F(2,40) = 7.70$, $p < .002$. Center frequency did not produce a significant effect, $F(2,40) = 1.80$, $p > .10$, nor was the interaction significant, $F(4,80) = 0.43$, $p > .10$.

Post-hoc testing with Tukey HSD (alpha = .05) revealed that the main effect of overlap was due to a significant difference in accuracy between the 0 octaves condition (65% overall accuracy) and the two lower overlap conditions. The -1 octaves condition, at 52% and the -2 octaves condition, at 50%, did not differ significantly from one another.

Overall performance was quite poor, at 55%. One-sample t-tests were performed against chance-level performance (50%) to determine which of the nine conditions was significantly above chance. The three 0 octaves conditions all were, with mean accuracies of 64%, 60% and 72% for center frequencies 5.3, 8.0 and 12.0, respectively ($t(20) = 2.828$, $p = .01$; $t(20) = 2.359$, $p = .03$; $t(20) = 3.697$, $p = .001$, respectively). All conditions with less than 0 octaves of overlap show chance level performance. The best score under these conditions was 56%, which was not above chance, $t(20) = 1.00$, $p = .32$.

Overall the data seem to suggest that for recognition to succeed at above chance levels, images must share some part of the spatial frequency spectrum. Images that are complimentary on the spatial frequency spectrum can be recognized from one another with limited success, but a gap between the images in terms of spatial content produces chance-level performance. This seems to hold equally true across the center frequencies tested here, though there may have been a trend towards better overall performance at higher center frequencies.

In comparison to Millward and O'Toole (1986), subjects seem to have done better in this experiment. With a center frequency of 11 c/o and an overlap of 0 octaves, they found 63% accuracy, whereas the most similar condition here (12 c/o center frequency, 0 octaves overlap) produced 72% performance. The difference may be due to this experiment using calibrated stimuli, whereas the earlier study used projected images whose luminance function was clearly non-linear (Millward & O'Toole, 1986). These non-linear images contain spurious spatial frequency content that might have hindered recognition. Although this interpretation is tempting, it must be viewed with some caution as there were a number of methodological differences between the two experiments that might also explain the results. Finding similar performance levels to Millward and O'Toole (1986) with a non-calibrated display would bolster this contention. This possibility was tested in Experiment 2a.

Though these results suggest that the threshold for above-chance recognition of faces is at or around 0 octaves under these circumstances, it is difficult to determine from them the degree of improvement in accuracy that results from a given increase in overlap and whether this will be similar at different center frequencies. This is due to the floor effect seen in the negative overlap conditions. In order to better assess the effects of varying overlap on performance, we examined higher levels of the factor in the following experiment.

Experiment 1b

This experiment extends the examination performed in Experiment 1a by looking at face recognition at higher overlap levels.

Method

Participants. Twenty-five undergraduates from McGill University (5 male, 20 female), with ages ranging from 19 to 44 years (median = 21), participated. All participants had normal or corrected-to-normal vision.

Materials. Materials were as described in the general stimulus description above. For this experiment, face images were processed by the filters shown in Figure 2b. Example stimuli may be seen in Figure 3b.

Design and procedure. The design and procedure of experiment 1b are exactly the same as experiment 1a, except that overlap levels of 1, 2 and 3 octaves were tested.

Results and Discussion.

Figure 4b shows the results for this experiment. The values given are mean accuracies with error bars representing one standard error. A two-way repeated measures ANOVA showed a significant main effect of overlap level ($F(2, 48) = 18.49$, $p < .00001$), a significant main effect of center frequency ($F(2, 48) = 3.63$, $p = .03$), and a significant interaction between these two factors ($F(4, 96) = 3.23$, $p = 02$). Overall performance was higher here than in Experiment 1, at 65%. This was as expected due to the higher levels of overlap being tested.

Post-hoc testing with Tukey HSD (alpha = .05) tests showed that the interaction was due to a non-significant simple main effect of overlap within center frequency 5.3 c/o. None of the overlap levels differed from one another at this center frequency. This is likely due to a floor effect, as all three levels of overlap at this center frequency produced performance near chance. At center frequency 8.0 c/o however, a clear effect of overlap is seen. The 1 octave overlap differed significantly from both 2 and 3, which did not differ from one another. Within center frequency 12.0 c/o, a similar tendency is seen, although here a significant difference is seen only between 1 and 3 octaves of overlap. Overlap levels of 2 octaves and 3 octaves did not differ significantly, likely due to a ceiling effect. The data suggest that higher center frequency produces superior results, as does higher spatial frequency overlap, but the latter factor tops-out at 2 octaves of overlap.

Conclusions from Experiment 1

Together, Experiments 1a and 1b show that overlap has a strong effect on face recognition performance, and that this effect combines with center frequency such that higher center frequencies produce better recognition at equivalent levels of overlap. In general, there seems to be an average rise of approximately 8.0% for every octave of increase in overlap, but this varies widely depending on center frequency. Each octave of change in center frequency produced an average increase in performance of 5.75%. This suggests that although both factors affect recognition, overlap has the more significant contribution under the conditions examined here.

In addition to determining the effects of overlap, we were interested to see what the effects of calibration vs. non-calibration were in this sort of paradigm. Non-calibration causes aberrant spatial frequency elements to be included in filtered images, which may hamper recognition. This should especially be the case when one is attempting to match two images with little overlap, because the extra spatial frequency elements introduced by the non-linearity would lower the correlation between the images, which is the only information available to the recognition process under these circumstances. We examined this question by also doing this experiment using a non-calibrated monitor in Experiment 2.

Experiment 2

This experiment is a replication of Experiment 1, using an uncalibrated monitor. The goal here was to determine how much of a difference in performance would occur when changing the monitor to a non-linear look-up table. This in turn allows suggestions to be made regarding the comparability of a number of past studies using non-calibrated filtered images as stimuli and more recent experiments that use a calibrated look-up table. The hypothesis is that effects will primarily be seen in conditions where there is low overlap. The reason for this is that under these circumstances, the only information available to the recognition process is the degree of correlation between the two images. No strict matching of same-frequency elements is possible. Adding in spurious elements to one image reduces the correlation between them and should therefore affect such recognition in a significant manner.

Experiment 2a

This experiment is the same as Experiment 1a, except that it uses an uncalibrated monitor to examine the magnitude of the effect on accuracy this manipulation might have.

Method

Participants. Twenty-four undergraduates from McGill University (12 male, 12 female), with ages ranging from 19 to 28 years (median = 24), participated. All participants had normal or corrected-to-normal vision.

Materials. Materials were as described in the general stimulus description above. Stimuli were the same as for Experiment 1a. For this experiment, face images were processed by the filters shown in Figure 2a. Example stimuli may be seen in Figure 3a.

Design and procedure. The design and procedure of experiment 2a are exactly the same as experiment 1a, except that the monitor used was not calibrated. Rather than using a linearized look-up table, the native 256 gray levels of the screen were used. The exponent of the monitor's gamma function of the monitor was approximately 2.4, which is fairly large (a typical value is around 2.2).

Results and Discussion.

Figure 5a shows the results from this experiment. The values given are mean accuracies, with error bars representing one standard error. A two-way repeated measures analysis of variance (ANOVA) showed no main effects of overlap, $F(2, 46) = 0.38$, $p = .69$, or of center frequency, $F(2, 46) = 1.45$, $p = .24$. There was no interaction either, $F(4, 92) = 0.05$, $p = .99$. The lack of main effects in this experiment is very likely due to the expected floor effect at lower overlap levels.

As with Experiment 1a, overall accuracy here was quite low at 56%. Although as a whole this was significantly better than chance ($t(23) = 3.40$, $p < .003$), one-group t-tests performed on each of the overlap by center frequency conditions showed that only the 0 octaves of overlap condition at center frequency 12.0 c/o was marginally greater than chance as an individual condition ($t(23) = 1.93$, $p < .07$). It is interesting to note that the accuracy level of this condition, at 60%, was consistent with Millward and O'Toole's findings under similar circumstances. For center frequency 11 c/o, overlap 0 octaves, and one practice session, they found 62% accuracy. The same condition using a calibrated

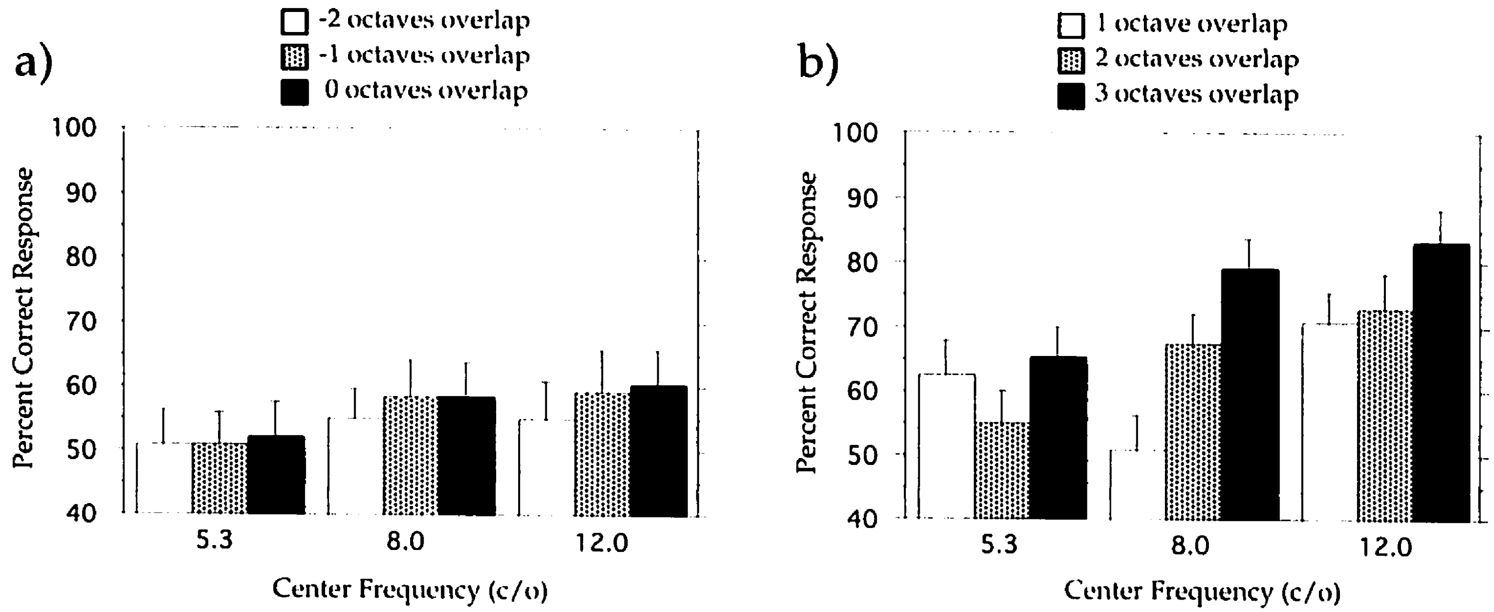

Figure 5. Mean accuracy data for Experiment 2. Error bars represent one standard error. a) Data for Experiment 2a. b) Data for Experiment 2b.

monitor (Experiment 1a) yielded a mean accuracy of 72% (Experiment 1a). This suggests that the low-frequency noise introduced into filtered images as a result of monitor non-calibration had an appreciable effect on recognition performance.

An overlap of 0 octaves at the other center frequencies produced chance-level performance, as did the lower overlap levels at all center frequencies. With a calibrated monitor, above-chance performance was achieved for 0 octaves of overlap at all center frequencies (Experiment 1a). This again suggests that monitor non-calibration has a deleterious effect on recognition performance. These results indicate that comparing studies using calibrated vs. non-calibrated stimuli can be somewhat difficult, as calibration can make the difference between chance and non-chance levels of performance. The fact that the lack of calibration seems to have had its greatest effect at the lower center frequencies is as predicted, as these are the conditions where the high-pass image has the broadest band and therefore there is the greatest potential for introduction of aberrant noise, although whether the signal-to-noise ratio is affected is difficult to ascertain.

To more formally explore the effect of calibration, a three-way mixed-design ANOVA was used to compare these findings with those of Experiment 1a. The design was 3 (overlap levels -2, -1, and 0 octaves) by 3 (center frequency 5.3, 8.0, and 12.0 c/o) by 2 (calibrated / uncalibrated monitor), with only the last factor being between subjects. The ANOVA showed an interaction between the calibration and overlap factors ($F(2, 86) = 3.21$, $p < .05$). Tukey HSD tests (alpha = .05) showed that this was due to the 0 octaves overlap conditions providing a greater advantage over other overlap conditions when stimuli were calibrated but not when they were uncalibrated. This further supports the assertion that calibration can make a significant difference in performance.

While the evidence thus far points to an appreciable effect of monitor calibration, the question remains as to whether this effect will only be seen when overlap is small. Under the conditions tested here, only the correlation between the two images is available to the recognition process. As overlap becomes greater, the effect of calibration may disappear as more direct information becomes available to the visual system for making the match. This possibility was examined by replicating Experiment 1b with a non-calibrated monitor.

Experiment 2b

This experiment is the same as Experiment 1b, except that it uses an uncalibrated monitor to determine if effects of calibration continue to be seen as overlap increases. Also, as with Experiments 1b and 1a, this experiment expands on the previous one by examining higher levels of overlap.

Method

Participants. Twenty-four undergraduates from McGill University (5 male, 19 female), with ages ranging from 17 to 39 years (median = 21), participated. All participants had normal or corrected-to-normal vision.

Materials. Materials were as described in the general stimulus description above and were the same as in Experiment 1b. For this experiment, face images were processed by the filters shown in Figure 2b. Example stimuli may be seen in Figure 3b.

Design and procedure. The design and procedure of experiment 2b are exactly the same as experiment 2a, except that higher levels of overlap were tested (1, 2 and 3 octaves of overlap). Stimuli were presented on an uncalibrated monitor.

Results and Discussion.

Figure 5b shows the results from this experiment. The values given are mean accuracy with the error bars representing one standard error. A two-way repeated measures ANOVA showed a significant main effect of overlap, $F(2, 46) = 7.71$, $p < .001$, and center frequency, $F(2, 46) = 6.68$, $p < .003$. The interaction between the two factors was marginally significant, $F(4, 92) = 3.02$, $p = .08$. Post-hoc testing with Tukey HSD (alpha = .05) showed that the main effect of overlap was due to the superiority of the 3 octave condition over the 1 octave condition. The difference between 1 octave and 2 octaves of overlap was small (4%) and non-significant. The difference between 2 octaves and 3 was significant, at

11%. This hints at a non-linear increase in performance with increasing overlap, but without an examination of finer gradations of this factor no firm conclusions can be made. Overall the results match those in the calibrated version of the experiment (Experiment 1b) quite well.

Similar post-hoc analyses showed that images at the 12.0 c/o center frequency were better recognized overall than the ones at the 8.0 and 5.3 c/o center frequencies, again matching experiment 1b in an overall manner. Post-hoc testing of the interaction between center frequency and overlap suggested that it was due to a greater advantage rendered by higher overlap levels at higher center frequencies. This was also similar to the findings of Experiment 1b.

Examining Figures 5a (results for Experiment 2a) and 5b (results for this experiment) shows an overall steady increase as overlap rises from -2 octaves to 3 octaves. The overall mean in this experiment was 68%, as compared to 56% in Experiment 2a. This increase in performance is quite similar to that seen between Experiments 1a and 1b. One-group t-tests revealed that most of the conditions in this experiment yielded above chance performance. The two exceptions were at the 2 octaves overlap condition at center frequency 5.3 c/o, which yielded 55% accuracy, and at 1 octaves overlap for the 12 c/o center frequency, where accuracy was only 51%.

As was done to compare Experiments 1a and 2a, a three-way ANOVA was used to compare the findings of the present experiment with those of Experiment 1b. The design was 3 (overlap levels -2, -1, and 0 octaves) by 3 (center frequency 5.3, 8.0, and 12.0 c/o) by 2 (calibrated / uncalibrated monitor), with only the last factor being between subjects. The ANOVA showed no significant overall effect of calibration, $F(1,47) = 0.704$, $p = .406$, and no significant interaction effects of calibration with either of the other factors (all $p > .40$). This supports the interpretations given above and shows that calibration did not seem to have a significant effect under circumstances of greater overlap. This is compatible with the prediction that calibration would generally have larger effects in cases where the learn and test images shared little or no spatial bandwidth.

Conclusions from Experiment 2

Taken together, Experiments 2a and 2b support the strong effect of overlap seen in Experiment 1, suggesting that this factor has an important role to play in explaining the human ability to match images in different display formats. Experiment 2 also replicates the finding that center frequency combines with overlap to determine overall accuracy in this task. Comparing Experiments 1 and 2 indicates that stimulus calibration has an significant effect on performance only in low-overlap situations. This is likely due to the fact that in this situation, only the correlation between images is available to drive the recognition process, whereas with higher levels of overlap recognition may be driven by more direct element-to-element matching. These results suggest caution when comparing older studies using slides or uncalibrated computer presentations as stimuli to more recent studies that typically use calibrated monitors, as the differences in calibration may produce significant differences in performance under certain circumstances.

Experiments 1 and 2 provide us with interesting information on how accuracy changes with spatial frequency overlap, but an alternative hypothesis for the results remains. There is a possibility that the effects observed arose in part or in whole from the change in bandwidth which accompanied changes in overlap. Two means of testing for this possibility are examined in subsequent experiments. In Experiment 3, conditions are tested wherein subjects must recognize images which are filtered in the same way as they are in Experiments 1 and 2, except that the recognition process is congruent. That is, subjects must match two images which have been filtered in the same way. This has the effect of holding overlap constant at 100% while changing the bandwidth of the images. If there is any effect of the bandwidth of the images, it should be seen under such circumstances. In Experiment

6, similar conditions to those in Experiments 1 and 2 are tested, except that the images are band-pass filtered to eliminate the problem of changing bandwidth.

In addition to its role as a control, Experiment 3 also provides baseline data for the previous experiments. By showing how recognition performance varies with the cut-off of low-passed or high-passed images, a basis for comparing the different levels of the overlap factor is derived.

Experiment 3

This experiment obtained baseline and control data for Experiments 1 and 2. Unlike in those experiments, where testing images were passed through filters opposite those of the learning images, in Experiment 3 learning and testing images are processed through the same filter. This has the effect of holding overlap constant at 100% while varying the bandwidth of the images, thus allowing the assessment of how much of the effects seen in the previous experiments was attributable to differences in bandwidth and filter cutoff. This in turn provides information on the utility of different regions of the spatial scale. The same face stimuli were used as in Experiment 1.

One aspect of the methodology for this experiment bears detailing here. In order to properly match the conditions in this experiment with those of Experiments 1 and 2, some of the filter levels were shown more than once. For example, the low pass filter with cutoff frequency 5.3 c/o was used three times in Experiments 1a and 2a: Once in the highest overlap conditions (-2 and 1 octaves respectively) at center frequency 5.3 c/o, a second time at the middle overlaps (-1 and 2 octaves respectively) at center frequency 8.0 c/o, and a third time at the lowest overlaps (0 and 3 octaves respectively) at center frequency 12.0 c/o. For this reason, this same filter was tested three times in the present experiment.

Experiment 3a

This experiment was designed to gather control and baseline data for Experiments 1a and 2a. It examined subjects' ability to match two images which were filtered in the same way, using as stimuli the same face images which were shown in those previous two experiments.

Method

Participants. Twenty-four undergraduates from McGill University (7 male, 17 female) with ages ranging from 19 to 40 years (median = 24), participated. All participants had normal or corrected-to-normal vision.

Materials. The face images used were the same as in Experiments 1a and 2a. They were presented on a non-calibrated monitor, matching conditions in Experiment 2.

Design and procedure. The experiment was a 2 (high-pass vs. low-pass filter) x 5 (filter ordinals 1 to 5) within-subject design. The relationship between this design and the 3 x 3 design of the Experiments 1 and 2 may be difficult to understand, but as the introduction to this experiment mentions, not all levels were presented the same number of times. Instead, images filtered with filter ordinal 3 were presented three times each, whereas those of filter ordinal 2 and 4 were presented twice each and those at filter ordinals 1 and 5 were presented once each. Adding these up shows the relationship between the 2 x 5 and 3 x 3 designs (i.e., 3 + 2 + 2 + 1 + 1 = 9 = 3 x 3). The experimental procedure was the same as for experiments 1a and 2a, except that faces at test were passed through the same filter as those at learning. That is, for every learned face, there was one target and one distractor in the testing set that had been filtered in the same way. This was true for the short initial practice session as well as the later experimental session.

Results and Discussion

As a first step in the analysis, subjects' correct responses were averaged across multiple presentations of given conditions (3 for ordinal 3 and 2 for ordinals 2 and 4). Figure 6a shows the results thus derived. Values shown are mean accuracy with error bars representing one standard error. Values are given for two filter types (high-pass vs. low-pass) at each of the five filter ordinals. Recollect that ordinals give an index as to the breadth of the band passed by the filter. Ordinal 1 represents the image with the broadest band whereas ordinal 5 represents the image with the narrowest spatial band. The associated cut-off frequencies can be seen in Figure 2a. Two cut-offs, one high-pass and one low-pass, are associated with each ordinal. The numbers at the tops of the bars in Figure 6a indicate the cutoff frequency of the associated filter.

Results were analyzed by means of a 2 (low-pass vs. high-pass) x 5 (filter ordinals) repeated measures ANOVA . This showed no significant difference between recognition of low-passed and high-passed images, $F(1, 23) = 0.04$, $p = .84$. Nor was there a significant effect of filter ordinal, $F(4, 92) = 1.22$, $p = .31$. The interaction between these factors was also non-significant, $F(4, 92) = 1.87$, $p = .12$.

In order to create baseline values for comparison with those obtained in the test conditions of Experiments 1a and 2a, nine derived values were obtained. This was done by taking the average of the accuracy in for the two filters—one high-pass, one low-pass—used in each condition of Experiments 1a and 2a. For example, a baseline value for the -1 octaves of overlap condition at center frequency 5.3 c/o (see Figure 1a, top panel, dashed lines) was obtained by taken the mean accuracy for low-pass images cut-off at 3.5 c/o and for high-pass images cut-off at 8.0 c/o. The resulting accuracies with associated standard errors are shown in Figure 7a. By comparing Figure 4a (Experiment 1a) or Figure 5a (Experiment 2a) to Figure 7a one can see that performance here is much better overall, with a mean overall accuracy of 83%. This supports the idea that the poor performance seen in Experiments 1a and 2a is due predominantly to a lack of transfer between filtered images and not to any difficulties associated with a given spatial band.

Figures 7a and 6a both present the data from this experiment, though in different ways. Both suggest that there is very little effect of the actual spatial content of the images on face recognition performance in congruent conditions. These figures contrast with those

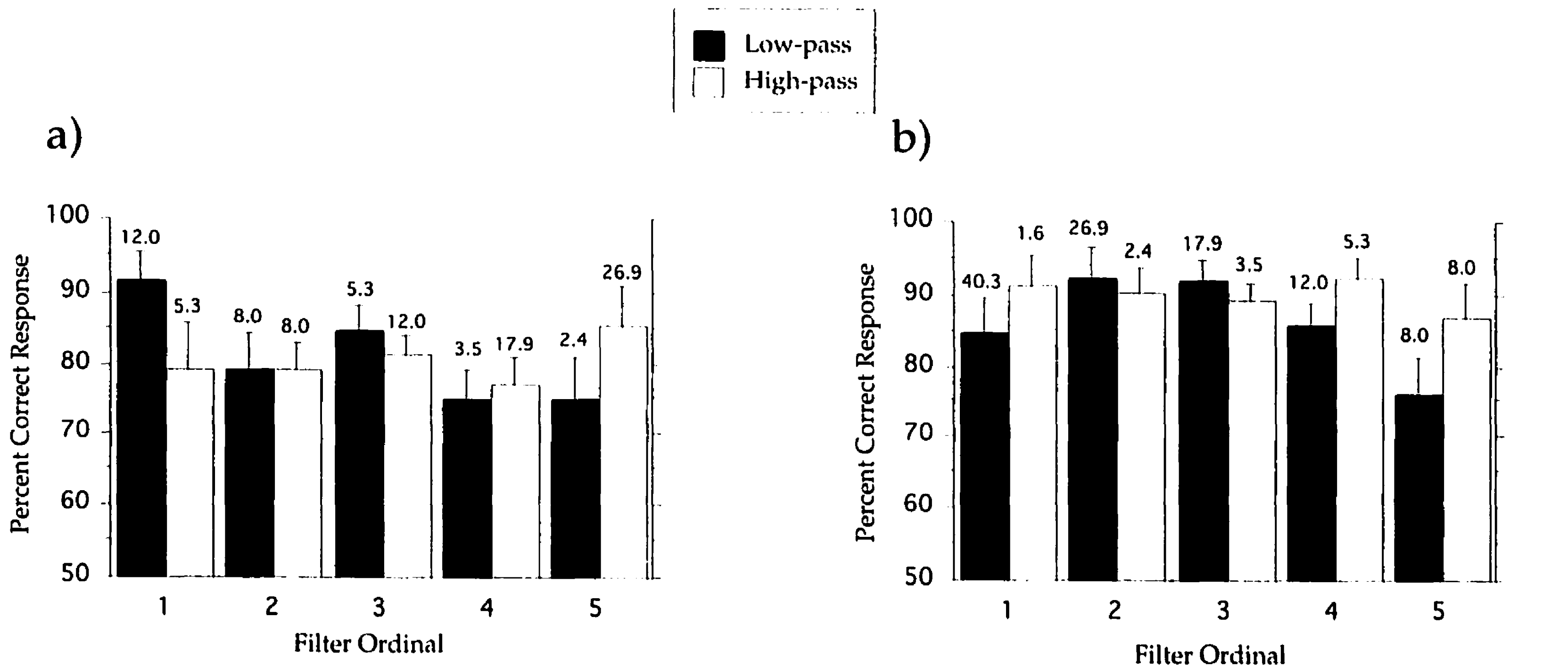

Figure 6. Control accuracy data for Experiment 3. Error bars represent one standard error. a) Data for Experiment 3a. b) Data for Experiment 3b.

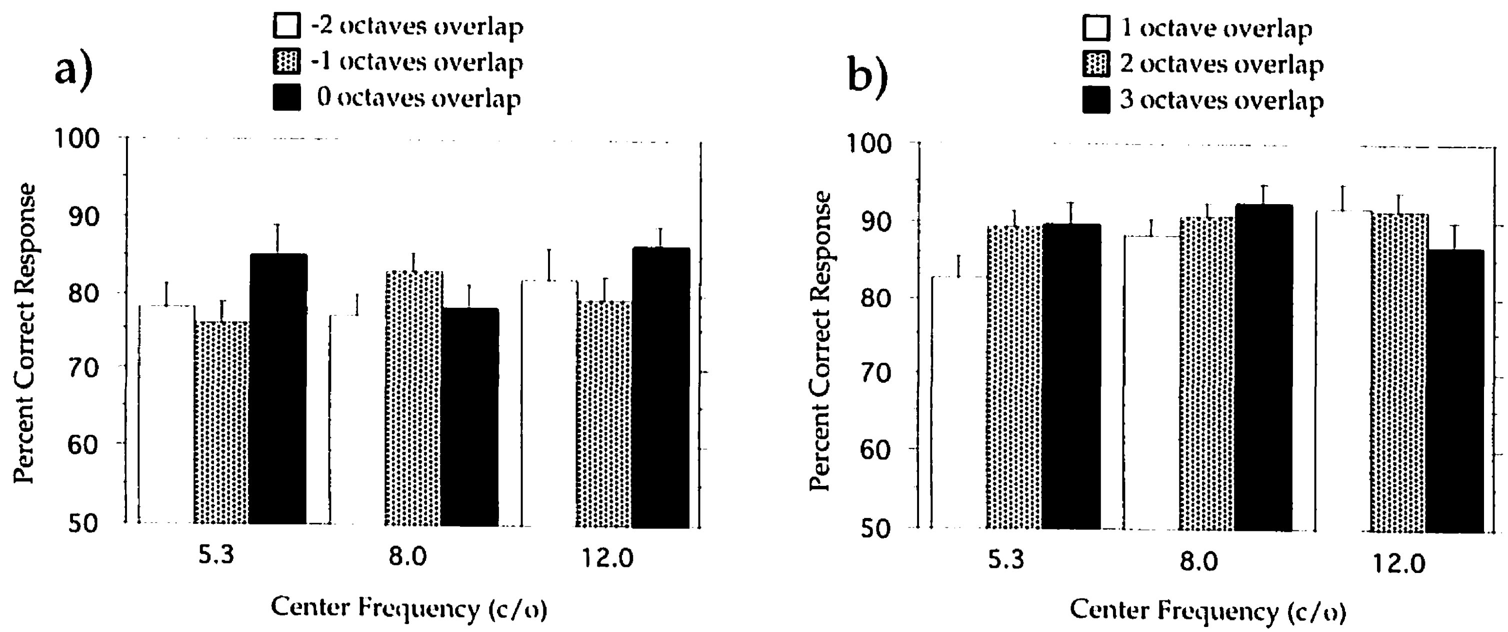

Figure 7. Baseline conditions derived from Experiment 3. Error bars represent one standard error. a) Data for Experiment 3a. b) Data for Experiment 3b.

from Experiments 1a and 2a, which both show a tendency for performance to improve with higher overlap and higher center frequency.

Experiment 3b

This experiment was designed to gather control and baseline data for Experiments 1b and 2b. It examined subjects' ability to match two images which were filtered in the same way, using as stimuli the same face images which were shown in those previous two experiments.

Method

Participants. Twenty-three undergraduates from McGill University (6 male, 17 female) with ages ranging from 19 to 47 years (median = 23), participated. All participants had normal or corrected-to-normal vision.

Materials. The face images used in this study were the same as those in Experiments 1b and 2b. They were presented on a non-calibrated monitor, matching conditions in Experiment 2.

Design and Procedure. The task and procedure were identical to Experiment 3a.

Results and discussion.

The data from this experiment were analyzed in the same manner as those from Experiment 3a. Figures 6b and 7b show the results. Values given are mean accuracies and associated standard errors. As with the previous experiment, the data are presented in two formats. Figure 6b shows them in a 2 (high-pass vs. low-pass filters) by 5 (filter ordinals 6 to 10) format while Figure 7b gives baseline conditions in a 3 (center frequencies) by 3 (overlap levels 1 to 3).

The data were analyzed by means of a 2 x 5 within subjects ANOVA. This showed no significant effect of filter type (low-pass vs. high-pass), $F(1, 22) = 2.07$, $p = .16$, and only a marginal effect of filter ordinal, $F(4, 88) = 2.18$, $p = .08$. The interaction between these two factors was also non-significant, $F(4, 88) = 1.26$, $p = .29$. These results are quite similar to those of Experiment 3a, showing little effect of the actual spatial frequency range per se on face recognition and supporting the contention that effects seen in Experiments 1b and 2b are predominantly due to changes in overlap and not in changes of the actual filter ranges.

As with Experiment 3a, the accuracy for low- and high-pass images in each filter pair was averaged to create baselines for comparison between this experiment and the test conditions examined in Experiments 1b and 2b. The resulting baselines are shown in Figure 7b. As with the comparison between Experiments 1a/2a and 3a, the difference here is striking. Overall performance here was 89%, contrasting sharply with the overall accuracy of 68% in Experiment 2b. Again, the flat performance seen here makes it clear that the effects seen in Experiments 1 and 2 were due to difficulties with information transfer between low-passed and high-passed images as opposed to difficulties with filtered images in general. Figures 4b and 5b (Experiments 1b and 2b) show that when the floor effect in Experiments 1a and 2a was eliminated, performance was clearly determined by the degree of overlap between learned and test images and by the center frequencies.

Conclusions from Experiment 3

Overall, the results of Experiment 3 support the assertion that the effects seen in Experiments 1 and 2 were indeed due to changes in overlap, and not to the expanding bandwidth of the images as overlap increased. Both the basic control values and the derived baselines showed fairly flat functions with only a small and non-significant effect of image bandwidth.

In contrast to many previous studies, Experiment 3 suggests that no particular band of spatial frequencies is more useful than any other for the task of face recognition. Indeed, the conclusion to be drawn from Experiments 1 to 3 is that spatial content per se has little effect on face recognition until images are severely degraded. There was no suggestion in these data of a critical band of frequencies used in face recognition. If there had been, we would have expected to see a hump-shaped performance function. Together, Experiments 1 to 3 suggest a strong effect of the similarity between spatially filtered images, with a lesser effect of the actual range of spatial content. In this light, previous results suggesting a critical band may be seen to be influenced by the degree of agreement in bandwidth between images.

Experiments 1 to 3 examined the effects of spatial frequency overlap in face recognition using a recognition paradigm. Sergent (1986) has argued that the type of experimental paradigm used can strongly affect how spatial frequencies are used by the visual system to perform recognition tasks. It may be that overlap will have different effects depending on whether subjects are attempting to recognize a new sensory input from the memory of a previous one, or attempting to match two readily available sensory inputs. In the following two experiments, the effects of spatial frequency overlap are examined in a simultaneous matching paradigm. In addition to examining whether task type will affect recognition, it allows the study of how spatial frequency overlap affects face perception when minimal memory load is imposed. This in turn should raise overall performance levels and thus allow an examination of how changes in overlap at lower levels (i.e., negative overlap levels) affect recognition.

Experiment 4

Although floor-level performance was observed at negative overlap levels in Experiments 1a and 2a, it is clear that the same levels of overlap under different circumstances could produce above-chance performance. For instance, if subjects are able to see two images simultaneously, this should eliminate the need for a long-term memory trace and subsequent decay or retrieval interference, thus enabling greater accuracy. If this is case, one may be able to see a smooth progression of performance across these lower overlap levels, as was seen for higher overlap levels in Experiments 1b and 2b.

This possibility was examined in this experiment. The same overlap and center frequency conditions as were tested in Experiments 1a and 1b were tested here in Experiments 4a and 4b, respectively. The goal of these experiments was to determine the minimum level of overlap needed for above-chance recognition under these circumstances, which presumably represent the easiest form of face identification task. Because it was likely that accuracy levels would reach ceiling, both percentage of correct responses and reaction time were measured.

The following experiment also replicates the baseline conditions of Experiment 3, except that in this case the control conditions are tested within subjects. That is, in addition to test trials in which subjects were exposed to two faces sharing limited bandwidth, subjects were also exposed to control trials in which the two images on the screen were processed with identical filters. This allowed us to examine the effects of simple congruent bandwidth in a matching task.

Experiment 4a

This experiment examined simultaneous matching of faces under the same overlap conditions as Experiments 1a, 2a and 3a.

Method

Participants. Nineteen undergraduates from McGill University (3 male, 16 female) with ages ranging from 17 to 26 years (median = 20), participated. All participants had normal or corrected-to-normal vision.

Materials. The face images used in this experiment were the same as in Experiments 1a, 2a and 3a.

Design and procedure. The experimental design was similar to that of previous experiments, except that the test conditions (3 overlap levels x 3 center frequencies) and control conditions (2 filter types x 5 filter ordinals) were evaluated within subjects.

The experiment began after subjects read the instructions on the monitor screen. At each trial, two face images were presented simultaneously on the screen, one above the other. The distance between the two images was 2 cm, or 9 cm center-to-center. The two images were either processed by opposite filters from a high-pass / low-pass filter pair (experimental conditions) or were processed by the same filters (control conditions). Vertical placement of the images was randomly determined so that all conditions were properly counterbalanced. The order of conditions was also randomized. Subjects pressed the space bar to start each trial. They were instructed to judge whether the two images on the screen were of the same face and to respond as quickly and accurately as possible.

Results and Discussion

Accuracy Data. The accuracy data for this experiment are shown in Figure 8a, which shows test condition data, and Figure 9a, which shows data from baseline conditions that were constructed in a similar manner to those in Experiment 3. The values given are mean accuracies with error bars representing one standard error. Separate analyses were run on the test and control data. A 3x3 completely within-subjects ANOVA for the test conditions showed a significant effect of both center frequency, $F(2,36) = 5.63$, $p = .007$, and spatial frequency overlap, $F(2,36) = 17.43$, $p < .00001$. The interaction was non-significant, $F(4,72) = .84$, $p = .50$.

Post hoc testing with Tukey HSD (alpha = .05) showed that the effect of center frequency was due to a significant difference between the 5.3 c/o conditions and the two higher conditions (8.0 and 12.0 c/o), which did not differ from one another. This replicates the previous findings of superior performance at higher center frequencies. Post-hoc testing was also performed to examine the effect of spatial frequency overlap. This showed that all three levels differed from one another, although the difference between the -1 and 0 octaves of overlap conditions was only marginally significant (p=.07).

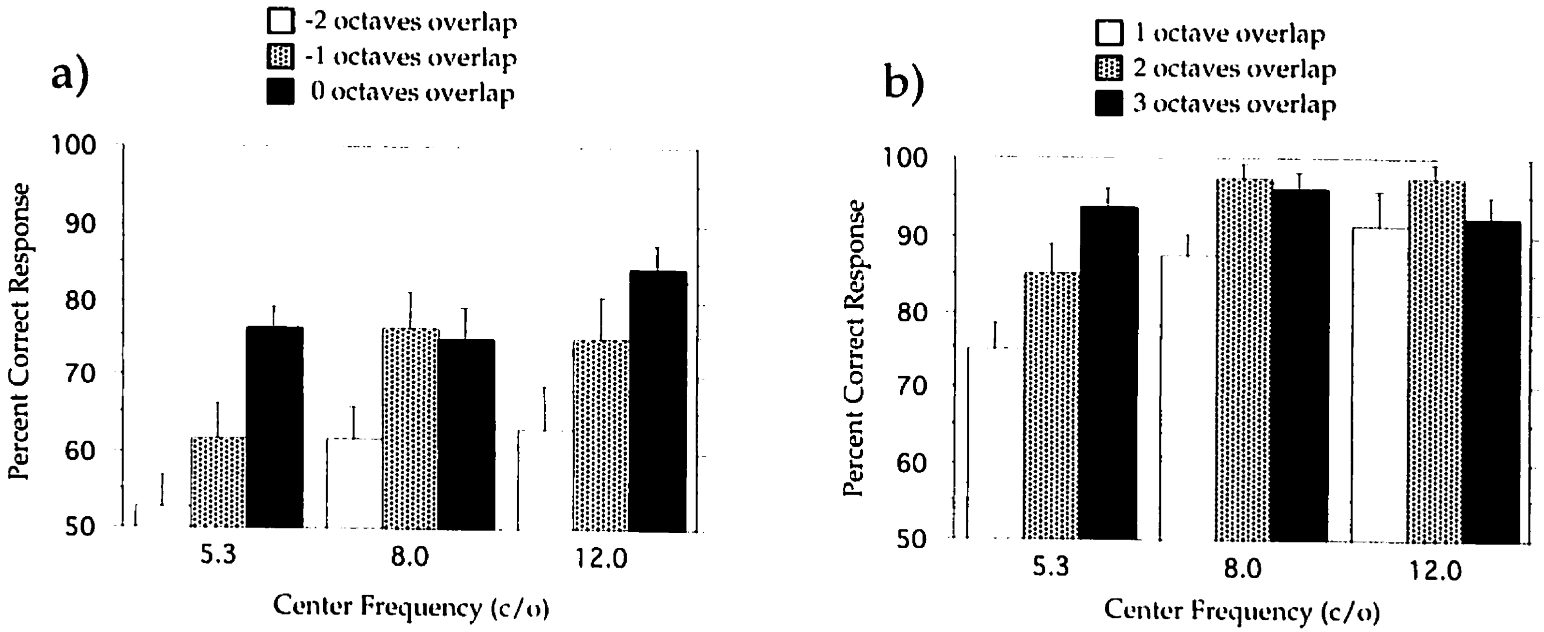

Figure 8. Mean accuracy data for Experiment 4. Error bars represent one standard error. a) Data for Experiment 4a. b) Data for Experiment 4b.

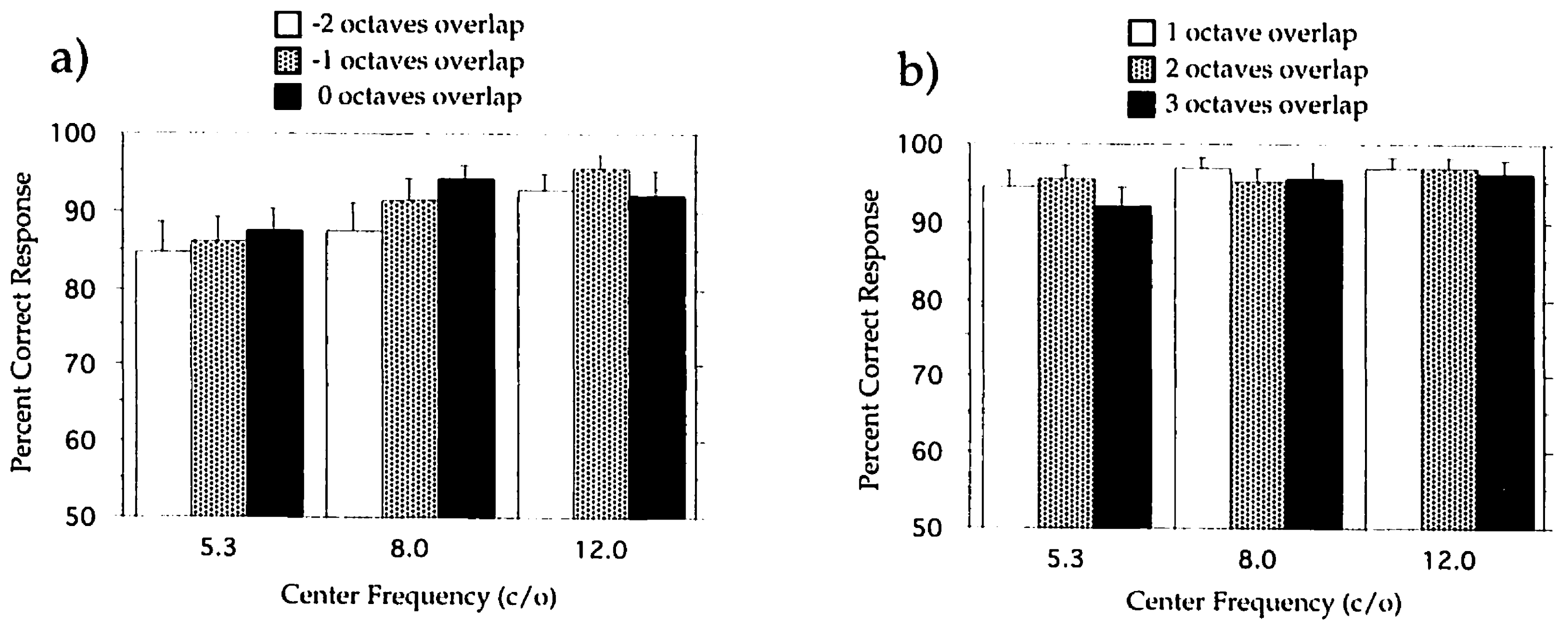

Figure 9. Mean baseline accuracy data for Experiment 4. Error bars represent one standard error. a) Data for Experiment 4a. b) Data for Experiment 4b.

Baseline conditions were subjected to a 2 x 5 ANOVA. This showed no significant effect of type of filter (high-pass vs. low-pass), $\underline{F}(1,18) = .79$, $\underline{p} = .39$, but did reveal a significant effect of filter ordinal, $\underline{F}(4, 72) = 7.15$, $\underline{p} < .0001$. The interaction was non-significant, $\underline{F}(4,72) = 1.31$, $\underline{p} = .27$. Post-hoc testing with Tukey HSD (alpha = .05) revealed that the effect of filter ordinal was primarily due to a difference between the narrowest filters (Ordinal 1) and the broadest ones (Ordinals 3 to 5). The latter did not differ amongst themselves, suggesting that very extreme filtering does have an effect on performance, but that increases in bandwidth past this point do not have an effect.

Reaction Time Data. The reaction time data for this experiment are shown in Figure 10a, which shows test condition data, and Figure 11a, which shows data from control conditions that were constructed in a similar manner to those in Experiment 3. The values given are mean accuracies with error bars representing one standard error. Separate analyses were run on the test and control data. A 3x3 completely within-subjects ANOVA for the test conditions showed a significant effect of spatial frequency overlap, $\underline{F}(2,36) = 3.66$, $\underline{p} = .04$. The effect of center frequency was not significant, $\underline{F}(2, 36) = 1.55$, $\underline{p} = .22$, nor was the interaction, $\underline{F}(4,72) = .68$, $\underline{p} = .60$. Post hoc analysis of the data show that the effect of spatial frequency overlap was due to superior performance in the 0 octaves condition as compared to the other two overlap conditions, which did not differ from one another.

Baseline conditions were subjected to a 2 x 5 ANOVA. This showed no significant effect of filter type (high-pass vs. low-pass), $\underline{F}(1, 18) = .57$, $\underline{p} = .46$, or filter ordinal, $\underline{F}(4,72) = .91$, $\underline{p} = .46$. Nor was there a significant interaction, $\underline{F}(4,72) = 1.73$, $\underline{p} = .15$.

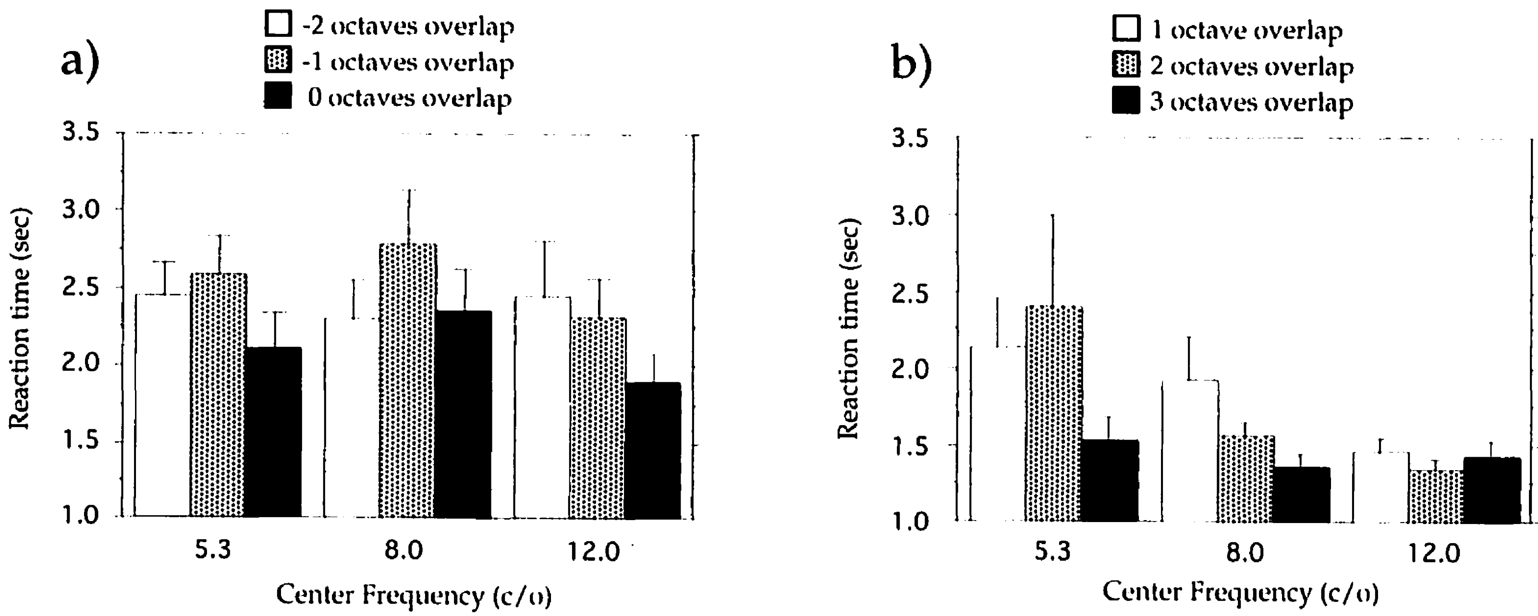

Figure 10. Mean reaction time data for test conditions in Experiment 4. Error bars represent one standard error. a) Data for Experiment 4a. b) Data for Experiment 4b.

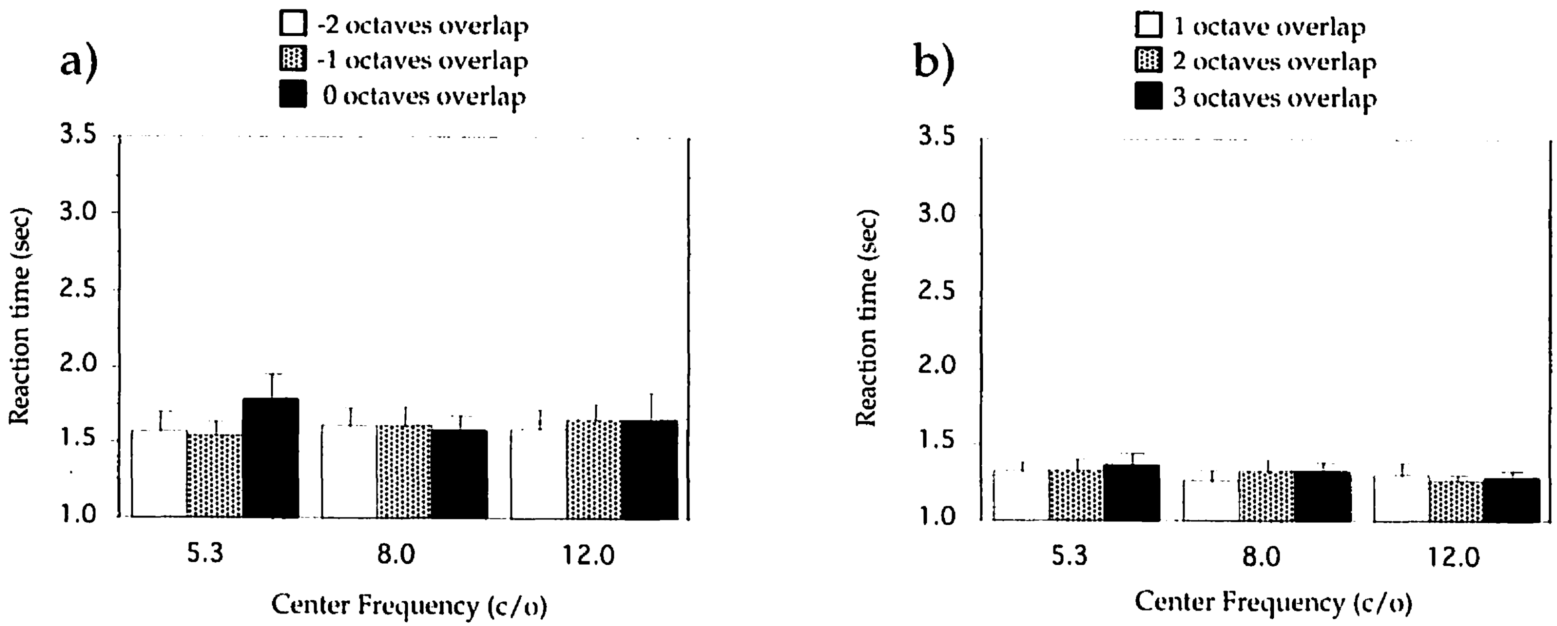

Figure 11. Mean reaction time data for baseline conditions derived for Experiment 4. Error bars represent one standard error. a) Data for Experiment 4a. b) Data for Experiment 4b.

Experiment 4b

This experiment examined simultaneous matching of faces under the same overlap conditions as Experiments 1b, 2b and 3b. It is similar in design and implementation to Experiment 4a, except that higher levels of overlap (1, 2 and 3 octaves) are tested.

Method

Participants. Twenty undergraduates from McGill University (4 male, 16 female) with ages ranging from 18 to 24 years (median = 20.5), participated. All participants had normal or corrected-to-normal vision.

Materials. The face images used in this study were the same as Experiments 1b, 2b, and 3b.

Design and procedure. These were the same as in Experiment 4a.

Results and Discussion

Accuracy data. The accuracy data for this experiment are shown in Figure 8b, which shows test condition data, and Figure 9b, which shows data from control conditions that were constructed in a similar manner to those in Experiment 3. The values given are mean accuracies with error bars representing one standard error. Separate analyses were run on the test and control data. A 3x3 completely within-subjects ANOVA for the test conditions showed a significant interaction between center frequency and overlap, $F(4,72) = 3.84$, $p = .007$. Post-hoc analysis with Tukey HSD (alpha = .05) revealed that this interaction was the result of overlap having a significant effect at center frequency 5.3 c/o but not at the higher center frequencies. An examination of Figure 8b shows that this is likely the result of accuracy levels reaching ceiling at the higher center frequencies.

Baseline conditions were subjected to a 2 x 5 ANOVA. This showed no significant effect of filter type, $F(1,18) = .02$, $p = .88$, but a marginally significant effect of filter level, $F(4,72) = 2.51$, $p = .05$. As with the previous experiment this was due to a significant difference between the narrowest filter level (Ordinal 1) and the broadest (Ordinal 5), again suggesting that only very extreme filtering had an effect on performance when overlap was 100%. The interaction was not significant, $F(4,72) = .38$, $p = .82$.

Reaction time data. The reaction time data for this experiment are shown in Figure 10b, which shows test condition data, and Figure 11b, which shows data from control conditions that were constructed in a similar manner to those in Experiment 3. The values given are mean accuracies with error bars representing one standard error. Separate analyses were run on the test and control data. A 3x3 completely within-subjects ANOVA for the test conditions showed a significant effect of both center frequency, $F(2,40) = 4.48$, $p = .02$, and spatial frequency overlap, $F(2,40) = 3.77$, $p = .03$. The interaction was not significant, $F(4,80) = 1.48$, $p = .22$.

Baseline conditions were subjected to a 2 x 5 ANOVA. This showed no significant effect of filter type, $F(1,20) = .01$, $p = .99$, or overlap, $F(4,80) = .77$, $p = .55$. The interaction was also non-significant, $F(4,80) = .59$, $p = .67$.

Conclusions from Experiment 4

As with Experiments 1 to 3, this experiment supports a strong effect of spatial frequency overlap. However, here we see that the effect generalizes to a different paradigm, confirming that such effects are not exclusive to matching images from memory. Results from this matching experiment are very similar to those from the previous recognition experiments (Experiments 1 to 3), indicating that there are not different effects based on task type.

Experiment 5

As was done with previous experiments, Experiment 4 was replicated with a non-calibrated monitor to determine what if any differences this would make to the effects observed.

Experiment 5a

This experiment was similar to Experiment 4a, except that stimuli were presented on a non-calibrated monitor.

Method

Participants. Twenty-four undergraduates from McGill University (9 male, 15 female) with ages ranging from 19 to 39 years (median = 22.5), participated. All participants had normal or corrected-to-normal vision.

Materials. The face images used in this experiment were the same as in Experiment 4a, except that they were presented on a non-calibrated monitor in this case.

Design and procedure. These were the same as for Experiment 4a.

Results and Discussion

Accuracy data. The accuracy data for this experiment are shown in Figure 12a, which shows test condition data, and Figure 13a, which shows data from control conditions that were constructed in a similar manner to those in Experiment 3. The values given are mean accuracies with error bars representing one standard error. Separate analyses were run on the test and control data. A 3x3 completely within-subjects ANOVA for the test conditions showed significant main effects of overlap and center frequency, $Fs(2, 46) = 21.09$ and 17.31, $ps < .0001$, respectively. The interaction was not significant, $F(4, 92) = 0.39$, $p = .82$. Post hoc testing with Tukey HSD (alpha = .05) tests revealed that the three levels of overlap each differed from one another significantly, with performance improving with increasing overlap. Similarly, the three levels of center frequency were each significantly different from one another, with increased accuracy accompanying higher center frequency.

Baseline conditions were subjected to a 2x5 ANOVA. This showed no effect of filter type (Low-pass vs. High-pass), $F(1, 23) = 0.51$, $p = .48$. The main effect of filter

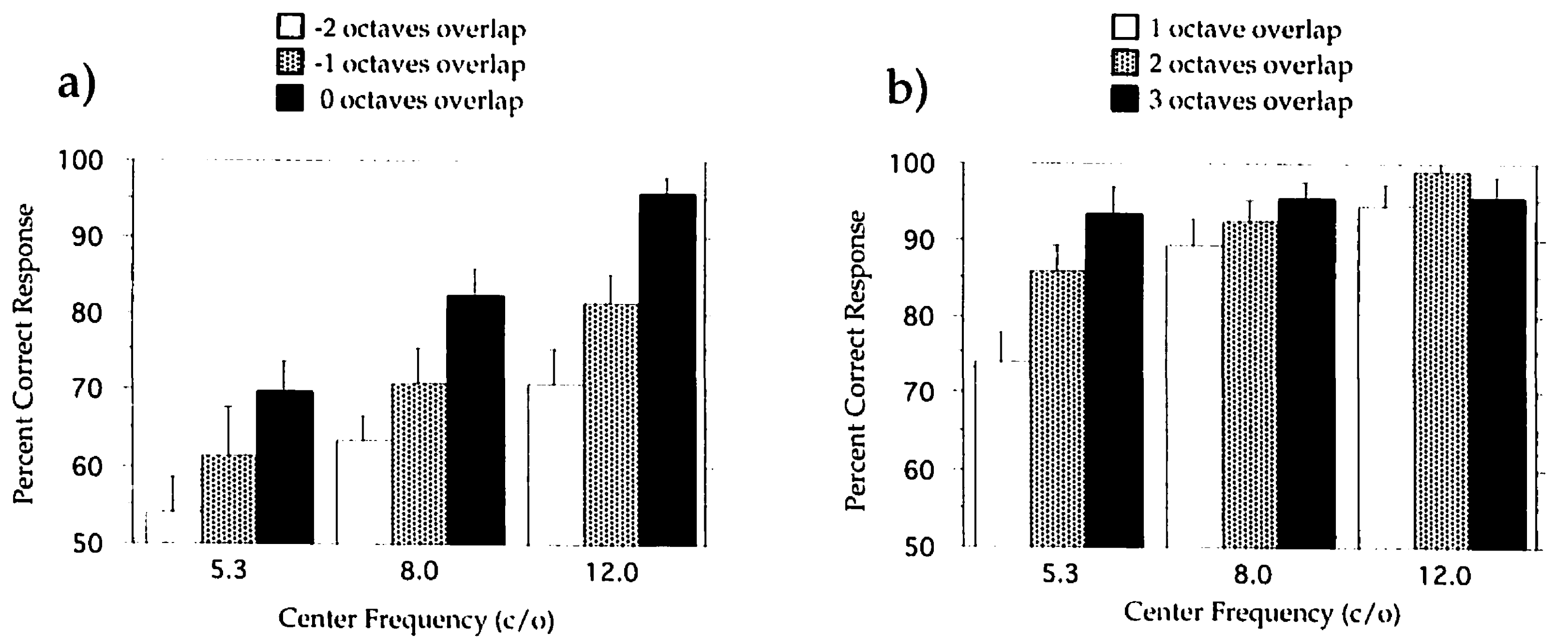

Figure 12. Mean test accuracy data for Experiment 5. Error bars represent one standard error. a) Data for Experiment 5a. b) Data for Experiment 5b.

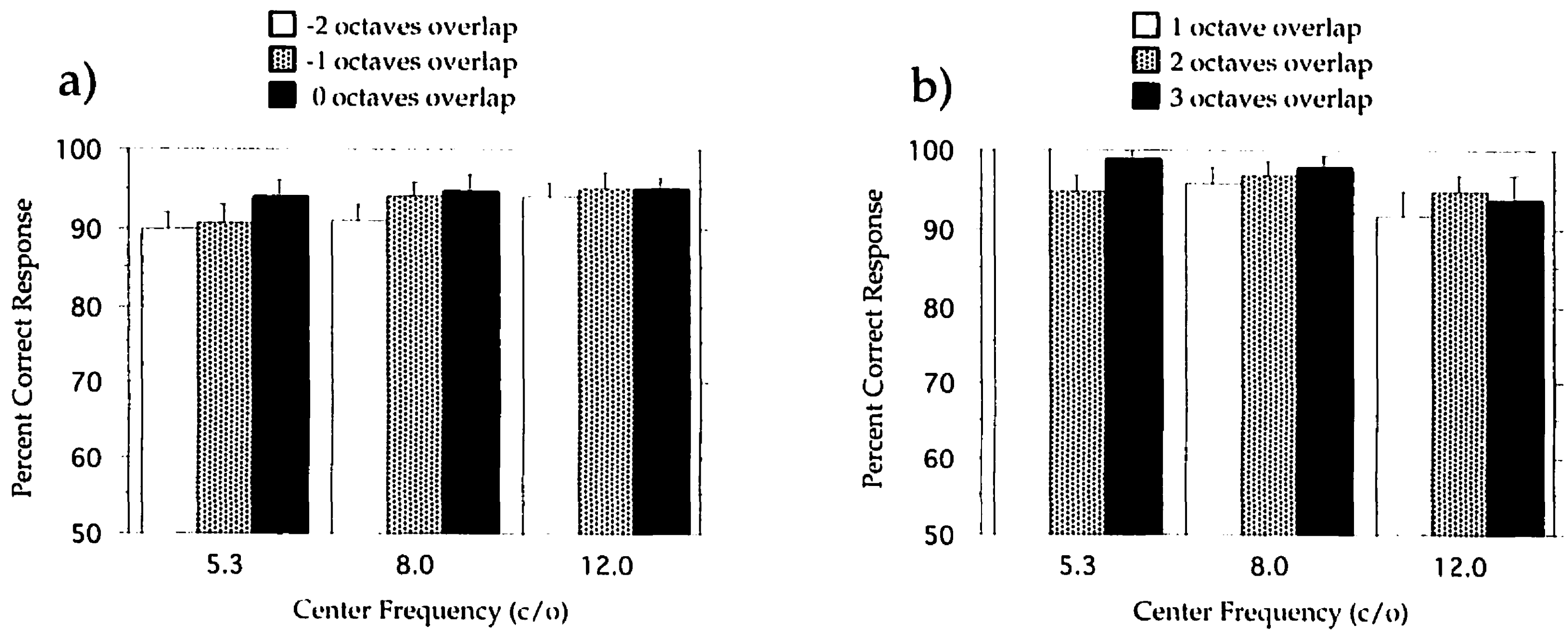

Figure 13. Mean baseline accuracy data for Experiment 5. Error bars represent one standard error. a) Data for Experiment 5a. b) Data for Experiment 5b.

ordinal was marginally significant, $F(4, 92) = 2.32$, $p = .06$, as was the interaction between filter ordinal and filter type (high-pass vs. low-pass), $F(4, 92) = 2.34$, $p = .06$. Figure 13a shows that this was due to a slight advantage of broad bandwidth images over narrower bandwidth images in the low-passed image conditions.

As was done to compare Experiments 1 and 2, a three-way ANOVA was used to compare the test data gathered using an uncalibrated monitor in this experiment to the data collected with a calibrated monitor in Experiment 4a. This revealed only significant effects of overlap, $F(2,82) = 20.5$, $p < .001$ and center frequency, $F(2,82) = 37.4$, $p < .001$. There was no main effect of calibration, $F(1,41) = .95$, $p = .336$, nor were any interactions significant.

A separate three-way ANOVA was applied to the control data in a similar fashion to the test data. This revealed a significant interaction between filter type and filter level, $F(4, 164) = 2.899$, $p = .02$, similar to that found in the individual analyses of the calibrated and uncalibrated data. The overall effect of calibration was non-significant, $F(1,41) = 2.76$, $p > .10$ as were all interactions with it (all $p > .10$).

Reaction time data. The reaction time data for this experiment are shown in Figure 14a, which shows test condition data, and Figure 15a, which shows data from control conditions that were constructed in a similar manner to those in Experiment 3. The values given are mean accuracies with error bars representing one standard error. Because the overall results match those of accuracy, with lower reaction times accompanying greater overlap and higher center frequency, the formal analysis of the data is not given here. Details are available from the author upon request.

Overall, the baseline conditions produced much faster reaction times (mean = 1.55 s) than the test conditions (mean = 2.44 s). The best reaction time for a test condition was at the highest center frequency and highest overlap, but this was still much slower than the corresponding baseline condition. Mean reaction time ranged wide across the test

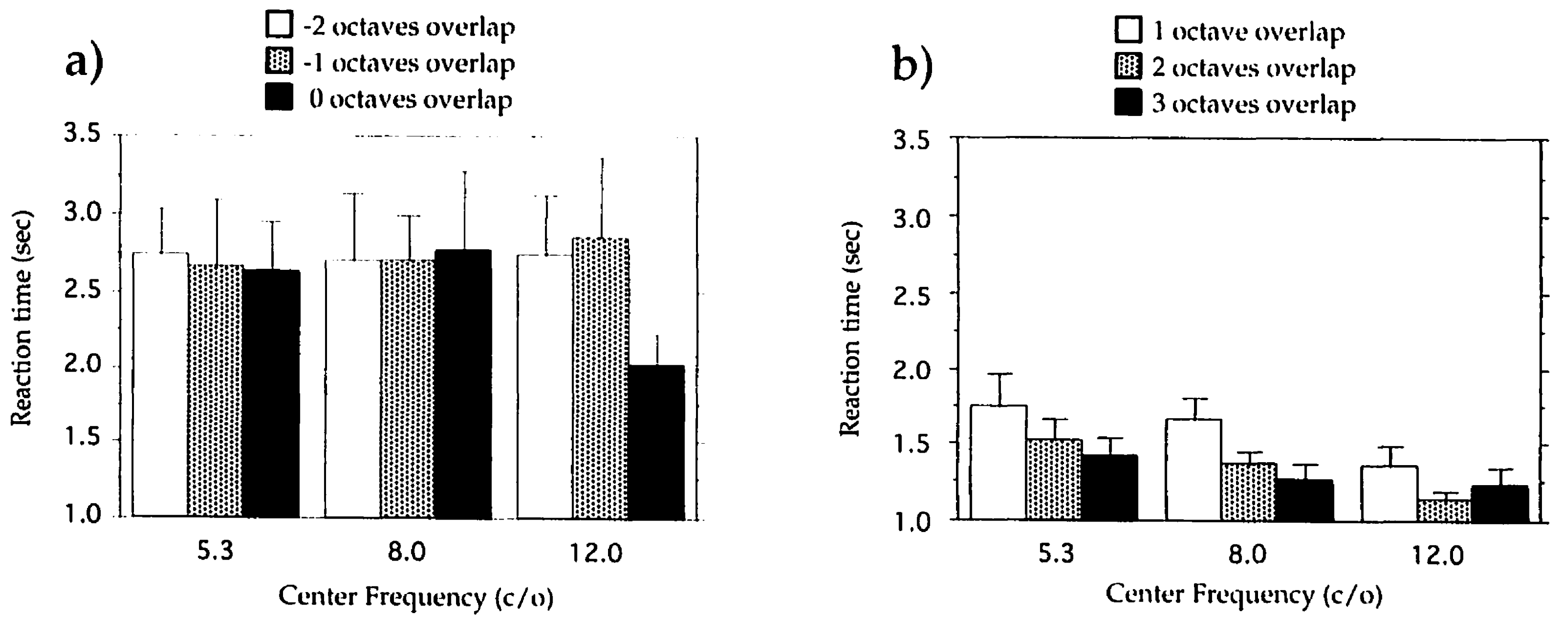

Figure 14. Mean reaction time data for test conditions in Experiment 5. Error bars represent one standard error. a) Data for Experiment 5a. b) Data for Experiment 5b.

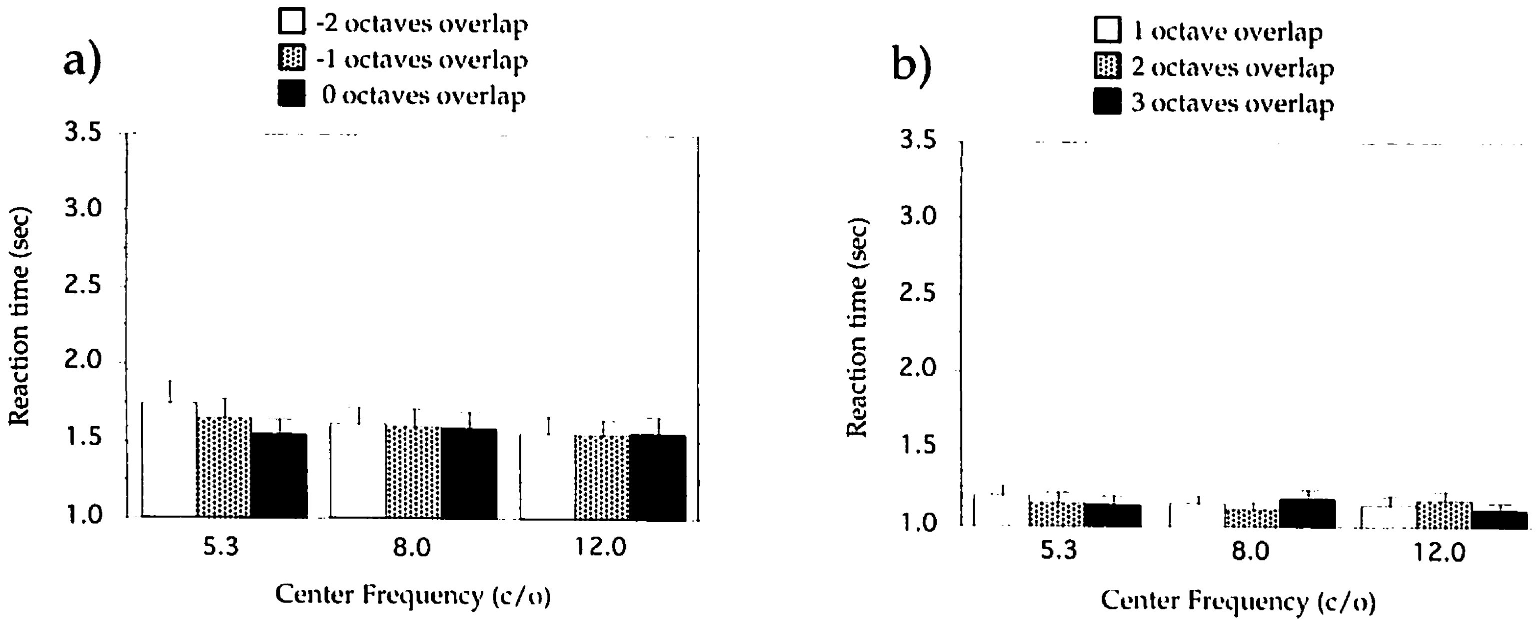

Figure 15. Mean reaction time data for baseline conditions in Experiment 5. Error bars represent one standard error. a) Data for Experiment 5a. b) Data for Experiment 5b.

conditions, from 1.86 s to 3.03 s. By comparison, the mean reaction times in control conditions were quite similar across conditions, varying from 1.44 s to 1.82 s. Again, this supports the contention that performance is much more affected by overlap than by the actual content of the images.

As with the previous experiments, we were interested to determine how performance would continue to improve as overlap increased above 0 octaves and at what point performance would reach ceiling. We examined this question in the following experiment.

Experiment 5b

This experiment was similar to Experiment 4b, except that stimuli were presented on a non-calibrated monitor.

Method

Participants. Twenty-four undergraduates from McGill University (5 male, 19 female) with ages ranging from 18 to 39 years (median = 20), participated. All participants had normal or corrected-to-normal vision.

Materials. The face images used in this experiment were the same as in Experiment 4b, except that they were presented on a non-calibrated monitor in this case.

Design and procedure. These were the same as for Experiment 4b.

Results and Discussion

Accuracy data. The accuracy data for this experiment are shown in Figure 12b, which shows test condition data, and Figure 13b, which shows data from control conditions that were constructed in a similar manner to those in Experiment 3. The values given are mean accuracies with error bars representing one standard error. Separate analyses were run on the test and control data. A 3x3 completely within-subjects ANOVA for the test conditions showed significant effects of overlap, $F(2,46) = 8.76$, $p < .0005$, and center frequency, $F(2,46) = 11.64$, $p < .0001$. The interaction between these factors was significant, $F(4,92)$, $p < .005$.

Post hoc testing with Tukey HSD (alpha = .05) showed that face images at 2 and 3 octaves of overlap were matched better than those at 1 octave of overlap at center frequency 5.3 c/o. Similar differences at 8.0 and 12.0 c/o were not significant. As with some of the previous experiments, overlap is seen to have a different effect at different center frequencies. In this case, it seems a ceiling effect occurs at center frequencies 12.0 c/o.

In the control conditions, no difference in accuracy was found between the low-pass and high-pass conditions, $F(1, 23) = 1.55$, $p = .23$, or between filter levels, $F(4, 92) = 1.66$, $p = .17$. No interaction was found between the two factors, $F(4, 92) = 0.55$, $p = .70$.

As was done with previous experiments, the data gathered using an uncalibrated monitor were compared with those gathered using a calibrated monitor by means of two three-way ANOVA separately applied to control and test data. The ANOVA for the test data revealed a significant interaction between center frequency and overlap, $F(4, 168) = 6.675$, $p < .001$, similar to that seen in the individual data sets. The overall effect of calibration was non-significant, $F(1,42) = .107$, $p = .745$, as were its interactions with center frequency and overlap (all $p > .10$). These results suggest a strong agreement between the two data sets.

A similar analysis run on the control data showed a significant effect of filter level, $F(4, 168) = 4.12$, $p = .003$, similar to that seen in the individual data sets being compared. The overall effect of calibration was not significant, $F(1,42) = .25$, $p = .622$. None of the other effects were significant (all $p < .24$). This suggests a strong agreement between the data from this experiment and those from Experiment 4b.

Reaction time data. The reaction time data for this experiment are shown in Figure 14b, which shows test condition data, and Figure 15b, which shows data from control conditions that were constructed in a similar manner to those in Experiment 3. The values given are mean reaction time with error bars representing one standard error. Separate analyses were run on the test and control data. A 3x3 completely within-subjects ANOVA for the test conditions showed significant effects of overlap, $F(2,46) = 19.81$, $p < .0001$, and center frequency, $F(2,46) = 12.57$, $p < .0001$. The pattern of the data here is the opposite to that for accuracy. That is, as accuracy increases, reaction time drops. The interaction was non-significant, $F(4,92) = 1.24$, $p = .30$.

Post-hoc testing with Tukey HSD (alpha = .05) showed that subjects responded faster when matching faces in the 2 and 3 octaves of overlap conditions than in the 1 octave of overlap condition. There was no difference between the two higher overlap conditions. Similar analyses showed that faces were matched faster at center frequency 12.0 c/o than at the lower two center frequencies, which did not differ significantly from one another.

Although accuracy data showed a significant interaction between center frequency and overlap in test conditions, the reaction time data did not. This was as we anticipated. As accuracy reaches ceiling, reaction time measures continue to capture the effects of overlap. Whereas accuracy data showed differences based on overlap only at the 5.3 c/o center frequency conditions, reaction time showed such effects at all center frequencies.

The baseline data was subjected to a 2 x 5 completely within-subjects ANOVA. This revealed a minute but consistent advantage of 20 ms for high-pass faces over low-pass ones, $F(1,23) = 4.45$, $p < .05$. The main effect of filter level approached significance, $F(4,92) = 2.23$, $p = .07$. The interaction was significant as well, $F(4,92) = 2.86$, $p < .03$. These reaction time data suggest that the lack of an seen in the accuracy data was due to a ceiling effect.

Post hoc testing with Tukey HSD (alpha = .05) revealed that the subjects took longer to match the most severely filtered low-pass faces (8.0 c/o) than to match other low-pass faces. The exception to this were the low-pass faces filtered at 26.9 c/o. This was not significantly different from those filtered at 8.0 c/o. No difference was found between reaction times for the high-pass faces. Clearly, this differential effect of filter level on low and high-pass faces was the cause of the significant interaction. The differences between reaction time for different filter levels were consistently smaller than those for different levels of overlap and center frequency in the test conditions.

A three-way ANOVA was used to compare the reaction time results of this experiment with those of the uncalibrated version (Experiment 4b). This showed no overall effect of calibration, $F(1,43) = 1.52$, $p = .225$, nor were any of the interactions with it significant (all $p > .10$). This suggests a good overall agreement between the data gathered on calibrated and uncalibrated stimuli.

Conclusions from Experiment 5

The results from this experiment are in general agreement with those of Experiment 4. Contrary to a comparison between Experiments 1a and 2a, which supports the idea that calibration has significant effects on performance, a comparison between Experiment 4 and 5 does not. The reasons for this are discussed in the General Discussion.

Stimuli for Experiments 6 and 7

The stimuli for these experiments were band-pass filtered images of faces and chairs. The face images were obtained from a database of 3D laser-scanned head models. The models were created using a Cyberware (tm) laser scanner that records both surface shape and texture of 3D forms. By mapping the texture map onto the shape model, one can create an image of the faces from any angle. The 3D face database was provided courtesy of Nikolaus Troje. Further details may be obtained in Troje and Bulthoff (1996; 1998).

The object images were obtained by photographing a variety of chairs from several different local areas. Chair images were chosen to remain consistent with Biederman and Kalocsai (1997; Kalocsai & Biederman, 1998) and because they are a homogeneous object category. All faces and chairs were imaged from the forward and 3/4 views. With faces, this means simply that the 3D head model was placed at 0° and 45° to the virtual camera before an image was captured. With chairs, the front corners of the seat were placed equidistant from the camera for front shots and the right front and rear left corners (or vice versa) were lined up relative to the camera for the 3/4 shot. There were 30 individual faces used and the same number of chairs. From each of these sets, 10 items were selected for use in practice sessions. All images were converted to 256 gray-level format before filtering and equated for mean luminance and RMS contrast.

The spatial filtering of the images was done with MatLab 5.2 software for MacIntosh. To create the filtered versions, the original full-bandwidth images were convolved with a pair of smooth Butterworth filters (one high-pass and one low-pass), then inversely transformed into the spatial domain. Butterworth filters were used to avoid the ringing effects seen when images are filtered with abrupt spatial frequency cutoffs. Nonetheless, the functions were steep enough to provide good spatial frequency localization. The low-pass filters were defined by

$$\frac{1}{1+(r/c)^5}$$

and the high-pass filters by

$$\frac{1}{1+(c/r)^5}$$

where r is the component radius and c the cutoff radius. To create bandpass images, each original image was sequentially passed through a low-pass and then a high-pass filter. The bands for each image were 2 octaves wide.

Filter cutoffs were selected so as to cover a wide range between the fundamental frequency of the images (1 c/o) and their folding frequency (about 68 c/o). Three overlap conditions were each defined at two center frequencies (7.1 and 14.2 c/o) by placing the bands symmetrically about them. Figure 16 illustrates the gain profiles and cutoff values of all the filters used in Experiments 6 and 7. Also shown, at the top of each filter function, is the middle frequency of that filter. As overlap levels increase, the image bands approach each other in .67 octave steps. That is, at each increase in overlap the low-pass cut-off in each filter pair moves up .33 octaves as the high-pass cut-off moves down .33 octaves. This results in three overlap levels of 0.33, 1.00 and 1.66 octaves.

Figure 17 shows examples of both face and chair stimuli passed through the filters shown in Figure 16. The rows represent center frequency and the columns overlap levels. Each cell contains two images, with the lower frequency band on the right and the higher frequency band on the left. The term "opposite filters" is used throughout the experimental descriptions to refer to bandpass filters symmetrically positioned about a given center frequency. For instance, at center frequency 7.1 c/o in the 1 octave of overlap condition (Figure 16, upper panel, dashed lines) the lower band ranges from 5.0 to 20.1 c/o, while the higher band ranges from 10.0 to 40.0 c/o. These are referred to as opposite filters to one another.

All experiments were run using a MacIntosh G3/233. Stimuli were presented on a 21" AppleVision Monitor that was properly calibrated to give a linear luminance profile. Images were surrounded by a medium gray background that filled the screen. Instructions were given on the monitor.

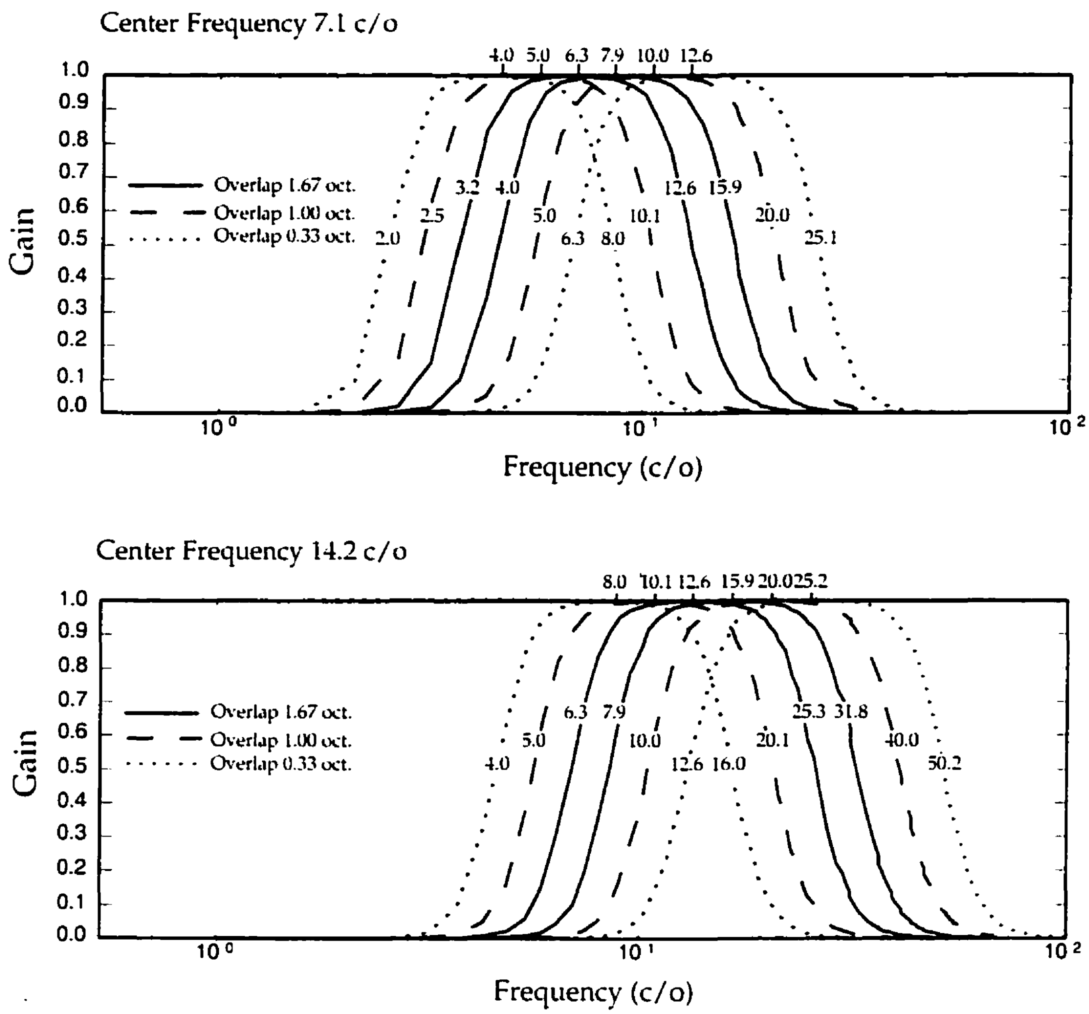

Figure 16. Gain profiles and 50% cut-off values of filters used to generate stimuli for Experiments 6 and 7. At the peak of each function is the center frequency of the filter.

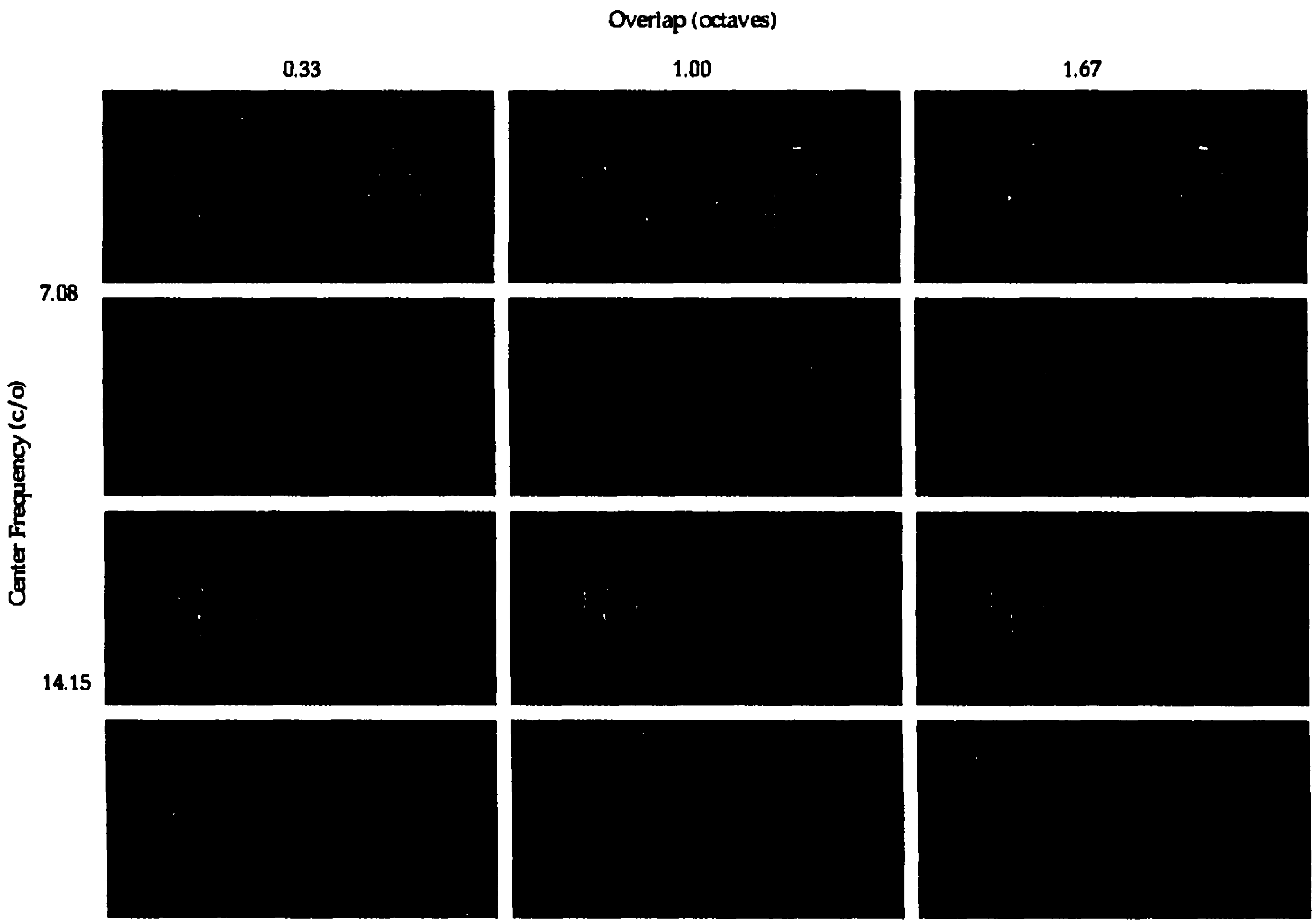

Figure 17. Example stimuli from Experiments 6 and 7. The image on the left is always from the lower band.

Experiment 6

This experiment was designed to determine if the spatial frequency overlap effects seen in Experiments 1 to 5 would generalize to object images. According to Biederman and Kalocsai (1997; Kalocsai & Biederman, 1998), objects are stored in a format that is spatial-frequency free and more robust to variations in spatial frequency than the format in which faces are stored. If this is the case, we should see overall better performance from object stimuli than face stimuli. Also, there should be an interaction between the type of stimulus and the spatial frequency overlap level, such that faces will show greater effects of overlap than objects. This interaction is important, because it is difficult to attribute a simple overall increase in performance to spatial frequency factors, as there might be other qualities of the stimuli that produce such a difference.

There are several differences between the methodology of this experiment and the previous ones. For instance, this experiment employs a sequential matching paradigm rather than a recognition or simultaneous matching paradigm. Sequential matching was chosen because this is more relevant to Biederman and Kalocsai's model of face and object storage. The reason is that Biederman's RBC model (which he supports as the model for object storage) is primarily put forward as the means by which "rapid and automatic" recognition is achieved (Biederman, 1987). Therefore, an old/new learning paradigm might not be a fair test of the model. Likewise, a simultaneous matching paradigm allows several comparisons between the objects to be matched, which might tend to make features such as the aspect ratio and lightness of stimuli more salient. These aspects are thought to be of secondary importance to object recognition in the RBC model (Biederman, 1987).

A second difference in this experiment is that the stimuli are rotated in depth by 45° between learn and test. This is done to ensure that we are in fact examining object and face recognition as opposed to more general image matching. Although image matching certainly seems to be an important part of face and object recognition, there is also an aspect of these processes that deals with the ability to recognize things under varying viewing conditions.

A third methodological difference between the following experiments and the previous ones is that control conditions are tested in the same experiment as the test conditions, allowing both sets of data to be explored in the same statistical analysis and therefore allowing more direct comparisons between them. Thus, Experiment 6 combines what has been analyzed as two separate studies in the previous experiments.

A final difference between this experiment and Experiments 1 to 5 is the use of bandpass filters to process stimuli. This was done to avoid a potential limitation of the previous experiments, whereby one might suggest that the results were the due to the expansion of the lower band rather than overlap per se.

Method

Participants. Sixty-four undergraduates from McGill University (16 male, 48 female) with ages ranging from 17 to 39 years (median = 20), participated. All participants had normal or corrected-to-normal vision.

Materials. The stimuli were face and object images prepared as discussed in the section entitled "Stimuli for Experiments 6 and 7", above. Example stimuli may be seen in Figure 17.

Design and procedure. The design of the experiment was a 2 x 2 x 2 x 3 mixed model. The factors were: 1) Stimulus type (face or chair), 2) Task type (test vs. control), 3) Center frequency (7.1 or 14.2 c/o), and 4) Spatial frequency overlap (0.33, 1.00, and 1.67 octaves). The first two factors were between subjects whereas the latter two were within subjects.

Subjects were tested using a Macintosh G3/233 computer with a 21" monitor. The monitor was properly calibrated to correct for its gamma function. Images were surrounded by a neutral gray background that filled the screen. Instructions were given in written form on the monitor and encouraged subjects to give their answers as quickly as possible.

The initial part of the experiment was a short practice session. On each trial of the practice session, subjects were shown two sequentially-presented images separated by a mask. The first image was presented for one second. This was followed by the mask, which consisted of a scrambled version of the first stimulus created by dividing the image into an 8 x 8 grid and randomly shuffling the elements. The mask was also presented for one second. This was followed by the testing image, which stayed on the screen until the subject's answer was given. The image pair to be matched consisted of pictures that had been filtered with opposite filters. For instance, if the first face was filtered to contain only frequencies between 2.0 and 8.0 c/o, the second face would be filtered to contain only frequencies between 6.3 and 25.1 c/o (see Figure 16, top panel, dotted lines). Twelve practice trials were given, two each of the 6 possible center frequency by overlap conditions. Half the subjects were tested with chair stimuli and half with face stimuli. Assignment to groups was random.

The experimental session immediately followed the practice session. The procedure here was the same as in the practice, except that each condition was tested 40 times, for a total of 240 trials. The breakdown of these trials is 3 (overlap levels) x 2 (center frequency) x 2 (target vs. distractor trial) x 20 (different faces and chairs). The order of conditions was randomized with the constraint that the same face could not appear in sequential trials. In each condition, half the trials presented a 0° view of the stimulus first, followed by a 45° view, whereas in the other half this order was reversed. Angle order was also randomized.

Following the presentation of each sequential image pair, subjects responded as to whether the images were pictures of the same object or same face. This was done by means of buttons marked "yes" and "no" on the computer keyboard. Subject accuracy and reaction time were automatically recorded by the computer program.

Results and Discussion

Accuracy data. Accuracy data is shown in Figure 18. The values given are means and standard errors. A four-way ANOVA was used to analyze the results. This showed a significant three-way interaction between the factors of stimulus type (face vs. chair), trial type (test vs. baseline) and overlap level (0.33, 1.00 and 1.67 octaves), $F(2, 120) = 3.2$, $p < .05$. Post-hoc analysis using Tukey HSD (alpha = .05) showed that this was due to a significant effect of overlap on face recognition in the test trials, an effect that was not evident in any of the other three conditions. Thus it appears that overlap has a strong effect on face recognition but not on object recognition.

In the test condition with faces as stimuli, the three overlap levels each differed from one another significantly at 67, 72 and 79% accuracy (for 0.33, 1.00 and 1.67 octaves of overlap respectively). In the equivalent baseline conditions little difference was seen between levels, with percentages of 77, 80 and 79% for the three levels in order. A similar analysis for objects showed that overlap had little effect on recognition here, with differences that were small and non-significant. Accuracy values in the test conditions were 67, 70 and 71% for the three overlap levels in order. These were not significantly different from one another. In the control conditions, a similarly flat function is seen, with associated values of 68, 71, and 70%. This pattern of results is as predicted by Biederman and Kalocsai's (1997; Kalocsai & Biederman, 1998) dual representation model of face and object recognition.

Center frequency had a significant main effect, $F(1,60) = 11.62$, $p < .002$, but did not interact with other factors. Overall, the center frequency 14.15 c/o showed slightly greater accuracy, at 74%, than center frequency 7.08 c/o, which yielded 71%. This is in agreement with Experiments 1 to 5, which found better performance at higher center frequencies. In those experiments, the possibility existed that these effects were due to the broadening of the band of frequencies in the low-passed image. In this experiment, no such alternate explanation is possible due to the fact that images were band-passed. One can therefore safely attribute higher performance to the greater amount of information in a given band of overlapping frequencies at higher points in the spectrum.

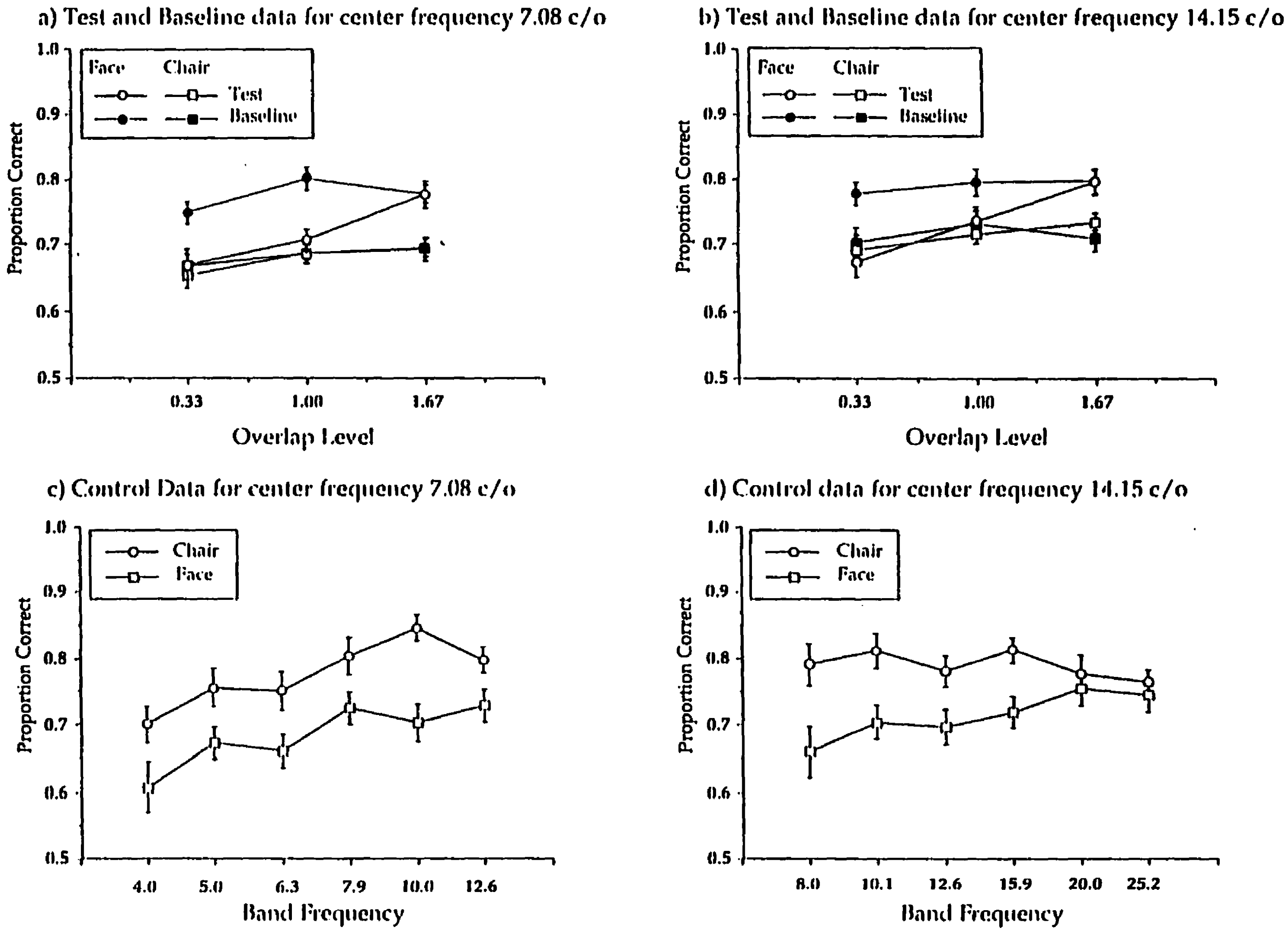

Figure 18. Mean accuracy data from Experiment 6. Test, baseline and control results are shown. Error bars represent one standard error.

Reaction time data. Reaction time data are shown in Figure 19. A four-way ANOVA was run to analyze this data. This showed no significant four-way interaction, nor were there any significant three-way interactions. Thus, the reaction time data failed to follow the accuracy data, assuaging concerns about a speed/accuracy trade-off. In general, the reaction time data were very similar across conditions, with the largest difference between conditions being about 60 ms. Despite this, a number of two-way interactions did prove statistically significant.

The interaction between trial type (test vs. control) and overlap (0.33, 1.00, and 1.67 octaves) was significant, $F(2, 120) = 12.48$, $p < .0001$. Post-hoc testing with Tukey HSD showed that this was due to a stronger effect of overlap in the test trials than in the control trials. In the former, mean reaction times were 1.22, 1.16 and 1.13 seconds in increasing order of overlap. These differences, on the order of 50 ms, are small but reliable and all levels differ significantly from one another. Similar values in the control conditions were virtually identical at 1.05, 1.03 and 1.02 seconds in the same order. As with the accuracy results, this suggests a strong effect of spatial frequency overlap with a weaker effect of actual spatial content of images.

The interaction between trial type and center frequency was also significant, $F(1, 60) = 7.59$, $p < .008$. Post hoc analysis with Tukey HSD showed that this was due to an effect of center frequency in test conditions -- where reaction time dropped from 1.19 to 1.14 seconds between center frequencies 7.08 and 14.15 c/o -- but not in control conditions, where the two conditions yielded virtually identical results at 1.04 and 1.06 seconds respectively. This seems to suggest that in matching images with 100% overlap, the actual content matters little, but when matching images with limited spatial frequency overlap, the higher frequencies provide more information with which to make the match.

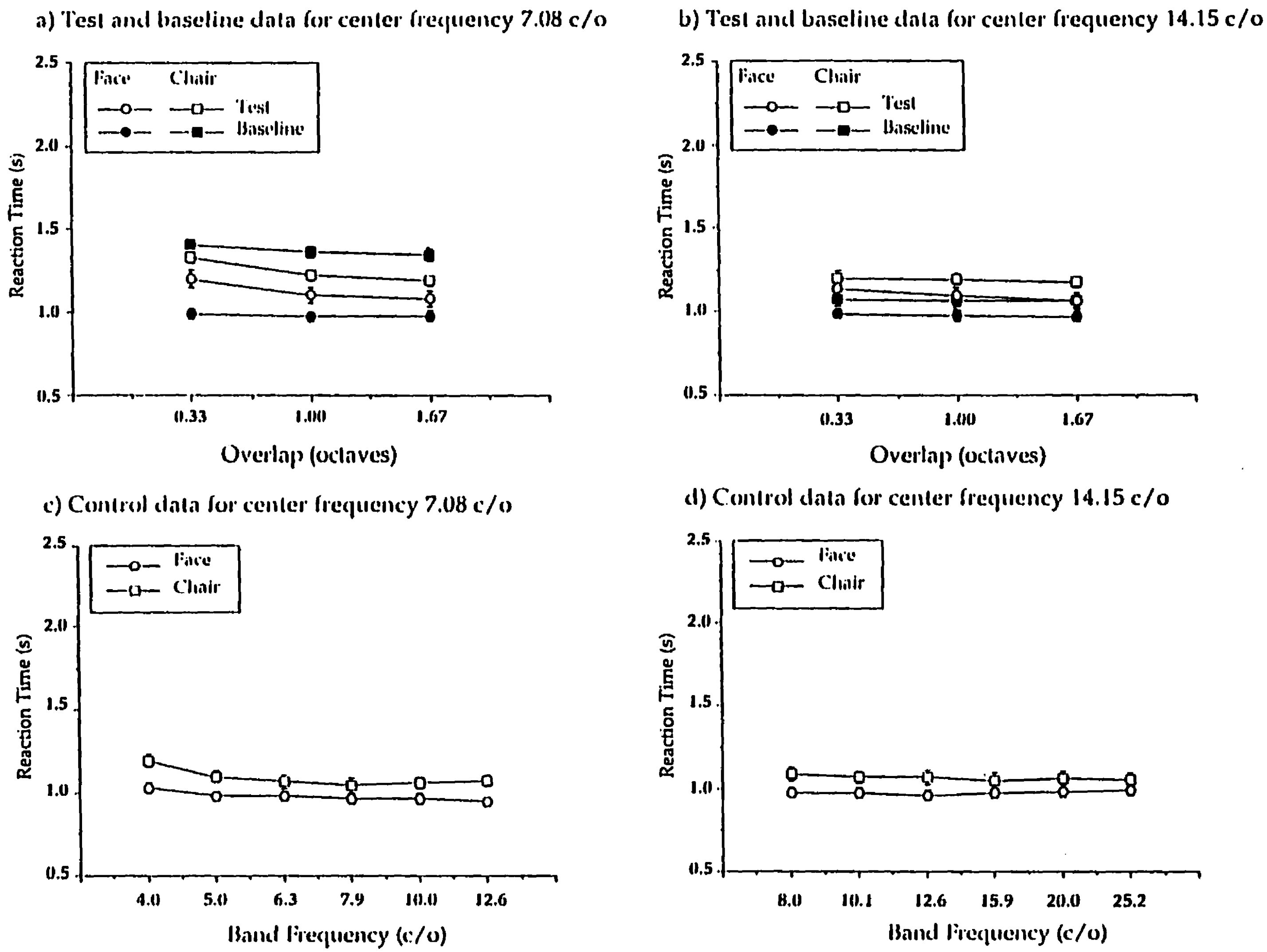

Figure 19. Mean reaction time data from Experiment 6. Test, baseline and control results are shown.Error bars represent one standard error.

A third significant two-way interaction was between stimulus type (chair vs. face) and center frequency, $F(1,60) = 5.95$, $p < .02$. This is due to a difference between center frequencies for chair stimuli--where reaction time dropped from 1.17 to 1.13 seconds--but not for face stimuli--where reaction times in the two center frequency conditions were virtually identical at 1.06 and 1.04 seconds respectively. This small but reliable difference suggests that higher spatial frequencies have an advantage in object recognition that is not as strong in face recognition.

A final two-way interaction was seen between center frequency and overlap factors, $F(2, 120) = 10.55$, $p < .0001$. In this case, reaction time dropped as overlap increased, but did so in different ways for the two center frequencies. At center frequency 7.08 c/o, overlap levels 0.33 and 1.00 octaves are significantly different from one another, but levels 1.00 and 1.67 are not. At center frequency 14.15 c/o, a more steady decrease is seen. Here levels 0.33 and 1.67 are significantly different, with level 1.00 in the middle being not significantly different from the other two. This result is likely due to the inclusion of faces with very low-frequency bandwidth in the lower center-frequencies 0.33 octaves of overlap condition. As Parker et al. (1996) have shown, low-band faces are recognized more slowly than middle or high-band ones.

Conclusions from Experiment 6

The accuracy data in Experiment 6 are clearly compatible with Biederman and Kalocsai's (1997) model of face and object recognition. Faces show a greater sensitivity to spatial frequency overlap and to spatial frequency content than objects do. Reaction time data was somewhat more difficult to interpret, and differences overall were quite small (on the order of 60 ms at most) but it did show that there were no accuracy for time trade-offs. Reaction time data also supported the effect of spatial frequency overlap being greater than that of spatial content in general.

Although the accuracy data can be seen as supporting a feature-based and spatial frequency-free representation of objects, there is an alternate explanation for object superiority in dealing with differences in spatial frequency content. That is, object images may correlate more strongly across spatial scale than do face images. Indeed, this seems likely, as object images typically contain more sharp edges, and edges tend to transcend spatial scale. If this is the explanation for the results of this experiment, then we should see a similar effect of spatial frequency overlap on the recognition of upside-down faces. These necessarily contain the same spatial frequency elements as upright faces, but may be treated more like objects by the visual system. If this is the case, we should see an effect of overlap on them that is somewhere between upright faces and objects. This possibility is examined in the following experiment.

Experiment 7

This experiment was designed to examine the alternative hypothesis that the effects seen in Experiment 6 were due a greater degree of correlation across spatial scale in the chair images than in the face images.

Method

Participants. Thirty-six undergraduates from McGill University (7 male, 29 female) with ages ranging from 16 to 28 years (median = 20), participated. All participants had normal or corrected-to-normal vision.

Materials. The stimuli were face and object images prepared as discussed in the section entitled Stimuli for Experiments 6 and 7. These were identical to the face stimuli used in Experiment 6, except that they were presented inverted. Example stimuli may be seen in Figure 17.

Design and procedure. The design of the experiment was a 2 x 2 x 3 mixed model. The factors were: 1) Task type (test vs. control), 2) Center frequency (7.1 or 14.2 c/o) and

3) Spatial frequency overlap (0.33, 1.00, and 1.67 octaves). The first factor was between subjects whereas the latter two were within subjects.

The procedure was identical to that used in Experiment 6, except that only face stimuli were presented, and these were presented inverted.

Results and Discussion

Accuracy data. Accuracy data is shown in Figure 20. The values given are means and standard errors. A three-way ANOVA was used to analyze the results. This showed a significant two-way interaction between trial type (test vs. baseline) and spatial frequency overlap (0.33, 1.00, and 1.67 octaves), $F(2, 68) = 7.74$, $p = .0009$. Post hoc testing with Tukey HSD (alpha = .05) showed that this was due to the overlap factor having a significant effect in the test conditions, where all levels were different from one another, but not in the control conditions, where none of the levels differed significantly. This result is similar to that found with upright faces, supporting an interpretation of the previous results based on differences in correlation across scale. None of the other interactions were significant, nor was the overall effect of center frequency.

Reaction time data. Reaction time data are shown in Figure 21. A three-way ANOVA was run to analyze this data. This showed a significant two-way interaction between center frequency (7.08 and 14.15 c/o) and overlap (0.33, 1.00, 1.67 octaves), $F(2, 68) = 4.19$, $p < .02$. Post hoc testing with Tukey HSD showed that this was due to an effect of overlap level at center frequency 7.08 c/o, but not at 14.15 c/o. This suggests that reaction time decreases as higher spatial frequencies are included in the task, reaching a floor level somewhere around the higher center frequency level. There was also an overall effect of trial type, $F(1,34) = 5.14$, $p < .03$, which showed that control conditions were overall slower. This is somewhat surprising, given that the test conditions would be expected to be more difficult. A possible explanation is that subjects were being more

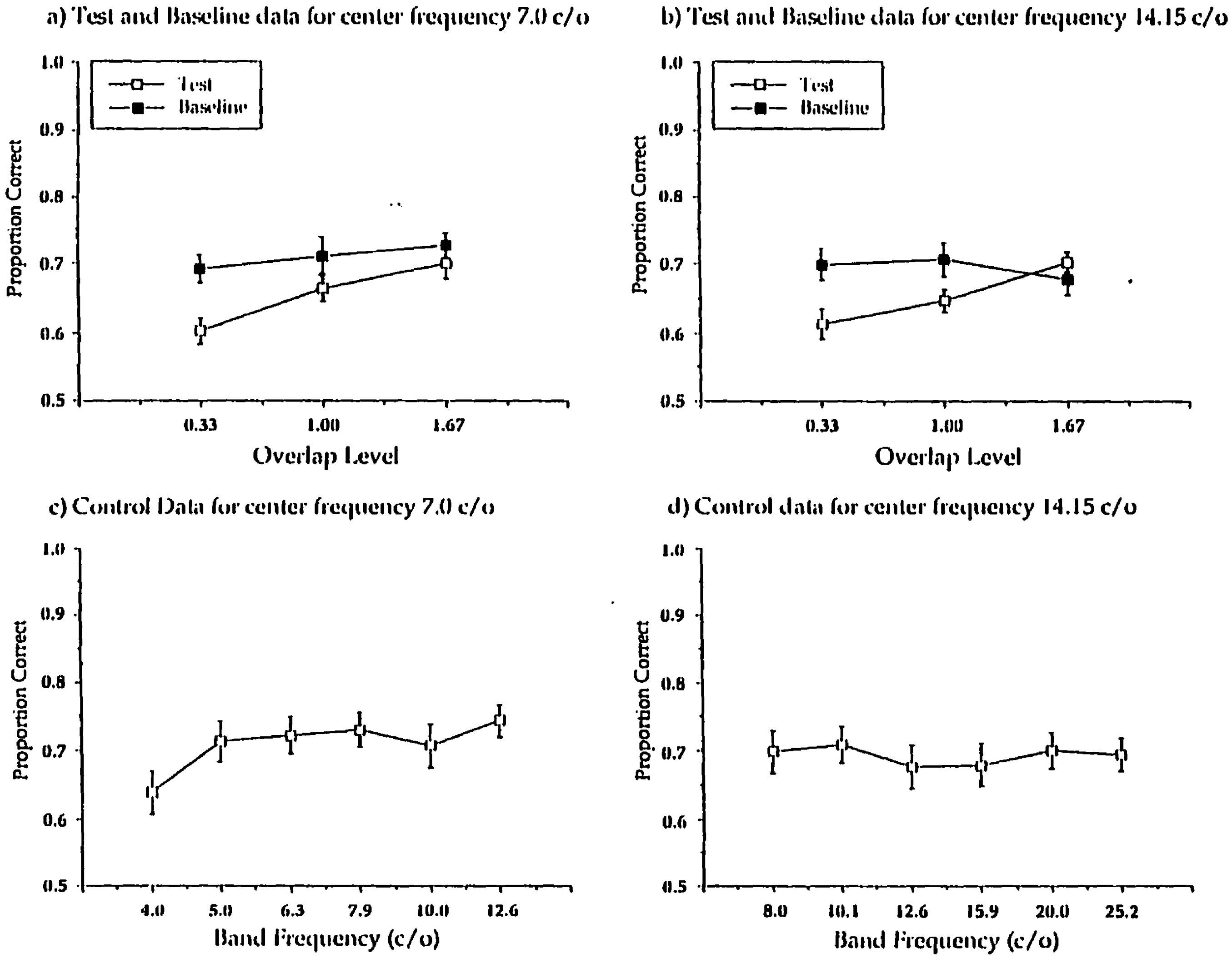

Figure 20. Mean accuracy data from Experiment 7. Test, baseline and control results are shown. Error bars represent one standard error.

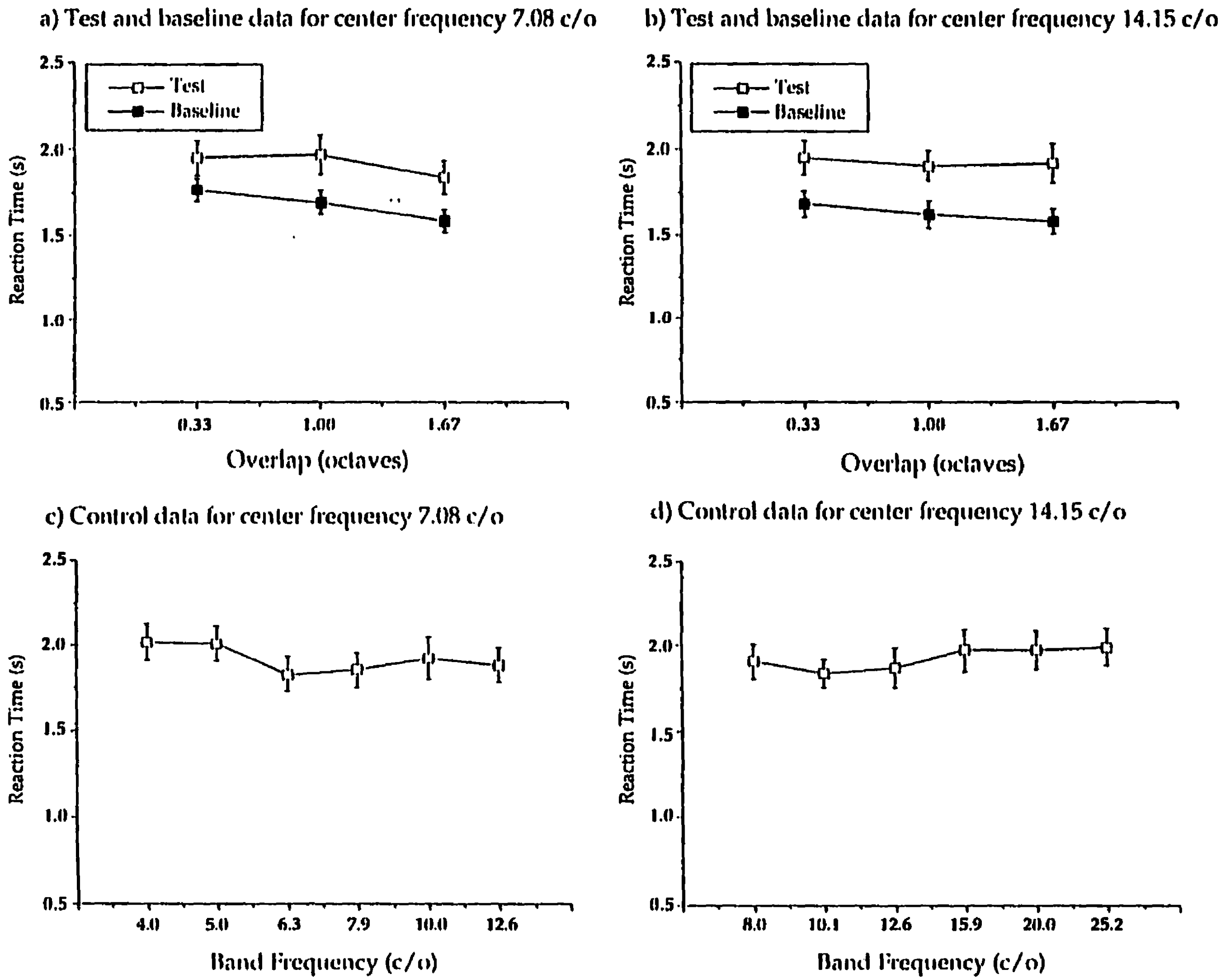

Figure 21. Mean reaction time data from Experiment 7. Test, baseline and control results are shown. Error bars represent one standard error.

cautious in the control conditions due to the fact that distractor stimuli would have a more similar overall appearance to the target stimuli in these conditions.

Conclusions from Experiment 7

The results of this experiment are compatible with the hypothesis that the greater resistance of object recognition to spatial frequency variations is due to a greater degree correlation across scale for object images. This explanation may in fact have broad applicability in explaining a number of findings which apparently support the RBC model of recognition (Biederman, 1987). This will be discussed in more detail in the General Discussion below.

CHAPTER 3: GENERAL DISCUSSION

This thesis examined the effects of spatial frequency overlap on face and object recognition, as well as a number of related questions. This was done in an attempt to provide a better understanding of how spatial frequency affects complex image recognition. In the following discussion, I first provide a brief review of the questions being addressed. This is followed by a detailed discussion of each question in turn. The discussion closes with some future directions suggested by the research and the general conclusions which may be drawn from this work.

The primary goal of this project was to assess whether spatial frequency overlap would have a significant effect on the recognition of complex images. Of particular interest was the relative magnitude of this effect as compared to the effects of simply varying the spatial frequency content of the stimuli. All of the experiments performed to examine this question found a significant effect of spatial frequency overlap and a relatively small effect of the band of frequencies to which a filtered image had been limited. These findings suggest that overlap between the spatial frequency domains of images is an important contributing factor to face and object recognition.

An additional question concerned the role of spatial frequency overlap in explaining higher-level phenomena such as difficulty with recognition across display format. The fact that overlap produced strong effects suggests that it may contribute a large portion of the effects seen in these cases. It may also help explain other phenomena, such as the effects of coarse-quantization.

Related to the magnitude of spatial frequency overlap effects is the question of which (if any) bands of spatial frequencies are most important for visual recognition tasks. The findings of the studies presented here suggest that the actual band of frequencies is of little importance, that a broad range of frequencies is sufficient for recognition, and that it is the similarity in spatial band between two images that is most important.

A second area discussed in the introduction and rationale concerned the effects of luminance calibration on recognition. If such effects are large, this would suggest that results from older studies using non-calibrated presentation methods could not be directly compared to modern studies that use calibrated stimuli. By executing experiments under both calibrated and uncalibrated conditions, it was found that this factor generally had little effect on recognition and matching tasks, although there was a significant improvement under certain circumstances. This suggests that the results of older studies will not be too different from those of modern studies, although some caution is warranted.

Another question examined in this thesis concerned the effects of task type. Of particular interest was the interaction of task type with spatial frequency effects. Sergent (1986) suggested that matching and recognition tasks might make preferential use of low and high frequency bands respectively. She argued that this explained some of the contradictory results in the literature. The studies reported here found similar effects of the spatial frequency factors under consideration in three different types of paradigms, and thus do not support Sergent's (1986) argument.

The final question posed in the rationale asked whether faces and objects differed in how variations in spatial frequency overlap affected them. This has implications regarding the format of the representation of these two types of stimuli in memory and specifically bears on Biederman and Kalocsai's (1997) model of face and object representation. The findings reported here support the notion that object recognition is less affected by the spatial frequency content of images, but do not support the notion that this is due to differences in how faces and objects are represented.

In the following discussion, each of the questions posed in the rationale will be examined in a separate subsection. Each subsection reviews the relevant results of this research and ties them in with past research.

<u>What is the effect of varying spatial frequency overlap?</u>

The main goal of this research was to determine the magnitude of the effects of spatial frequency overlap on recognition performance. Some degree of rise in performance with increasing overlap is to be expected based on the fact that the amount of information in common between the two images increases as they come to have greater overlap. However, the possibility exists that the effect of alterations in spatial frequency overlap would be small and therefore of little relevance to other questions in visual recognition. One reason to expect a small effect of overlap is that face and object recognition, being higher-level processes, might rely more on higher-order information and be little affected by the basic signal qualities of the input images. In performing face recognition, for example, the visual system may have access to a great deal of information about how faces "should be". At a simple level, this includes a weak constraint for bilateral symmetry and a strong constraint for the general organization of larger facial features. But it may also include such things as "eyes have sharp edges" and "cheek bones are never sharp edges". To the extent that the visual system has this sort of information available, it may be able to use it to over-come limitations in spatial-frequency overlap and various other sorts of band-width limitations. For example, a low-passed face image may have blurred lines around the eyes while a high-passed one may have a sharp line defining the cheek-bone. Innate knowledge about typical face features will facilitate recognition across these two representations by compensating for these "incorrect" portrayals of facial features. If this ability is very powerful, then one might expect very little effect of spatial frequency overlap. In this case, only extreme filtering would hamper recognition, either by making the image unrecognizable as face or by degrading the image so severely that insufficient information for making accurate extrapolations was provided.

If it was the case that effects of spatial frequency overlap were small, then this factor could be held to be of little relevance to other questions in visual recognition. But if effects are large, this argues that spatial frequency overlap must be taken into account as a contributor to higher-level phenomena. For instance, concerning the difficulty subjects have in recognizing line drawings of faces, a number of higher-level explanations have been proposed, such as a lack of mass representation (Bruce et al., 1992) and reduction in configural processing (Leder, 1996). While these likely contribute to the effect, it is difficult to know to what degree without first analyzing lower-level differences such as those in spatial frequency content. Similarly, the results of studies that examine the recognition of filtered images from unfiltered ones may be at least partially explained by the overlap factor if it is of sufficient magnitude.

Several paradigms were used to examine the effects of spatial frequency overlap on both face and object recognition. In general the results all showed that this factor had a strong effect on recognition, even when visual differences in stimuli were subjectively quite subtle. The results arising from each paradigm are discussed below, followed by a discussion of their implications.

In Experiments 1 to 3, a learn/test recognition paradigm was used with face images as stimuli to examine the effects of spatial frequency overlap. Here subjects were shown a series of learning faces in one session, and were tested to see if they could pick out these learned faces from a group of distractors in a second session. Experiments 1 and 2 examined subjects' ability to recognize images that shared a limited range of spatial frequencies, whereas Experiment 3 studied their capacity to recognize images that shared their whole range of spatial frequencies, but that varied in the portion of the spectrum they occupied.

In Experiments 1 and 2 it was found that variations in overlap could produce a full range of accuracy results, from floor to ceiling, and that subjectively minor differences in image appearance could produce significant differences in recognition rates. Experiments 1a and 2a looked at very low levels of overlap, ranging from -2 to 0 octaves. Here it was found that only images sharing some portion of the spatial spectrum could be matched at above-chance levels. That is, only in the 0 octaves overlap conditions was recognition significantly better than chance. In both Experiments 1a and 2a, recognition at negative overlap levels failed to be significantly better than chance at all center frequencies and under both luminance-calibrated and non-calibrated conditions.

Experiments 1b and 2b examined recognition with higher levels of overlap, from 1 to 3 octaves. In general, there was a steady rise in accuracy as level of overlap increased. It was found that the effect of overlap interacted with that of center frequency, such that higher center frequencies produced better recognition for a given level of overlap. Recognition reached ceiling levels around 2 octaves of overlap in most cases.

An interesting aspect of the findings of Experiments 1 and 2 is that subjectively minor changes in image appearance can produce significant differences in accuracy scores. Experiments 1a and 2a examined the recognition of pairs of images similar to those shown in Figure 3a, while Experiments 1b and 2b did the same for images like those in Figure 3b. In examining these figures, one would not expect such large differences based on the appearance of the various pairs of pictures. This suggests caution in attributing similarity to images of similar format (for instance two types of line drawings) that appear subjectively comparable, and once again argues that spatial frequency overlap may be a significant contributing factor to the effects seen in studies of display format.

Experiment 3 examined the ability of subjects to match two identically filtered face images. This experiment was designed to determine if the effects seen in Experiments 1 and 2 were indeed due to differences in overlap, or if they were due to decreases in the bandwidth retained by images as overlap decreased. That is, as overlap level dropped in the previous two experiments, the cut-offs of the filters through which learned and tested images were passed concomitantly became more extreme. Thus the images in lower overlap conditions had narrower bandwidths. If the lowering of bandwidth alone produced significant differences in recognition, this would suggest that effects attributed to overlap in Experiments 1 and 2 were in fact due to reductions in information presented in the learned and tested images. Note that because images filtered in the same way were using at learning and testing in these conditions, overlap in spatial frequency was held constant at 100%. Only the spatial frequency band of the images to be recognized was changed.

A second aim of Experiment 3 was to produce baseline data for the previous two studies. Baseline data for each of the 9 test conditions (3 overlap by 3 center frequency) in Experiments 1 and 2 was derived by averaging the accuracy scores from two conditions in Experiment 3. In Experiments 1 and 2, learn and test images were put through opposite filters, one high-pass and one low-pass. In Experiment 3, the baseline conditions were derived by averaging the accuracy for matching the associated high-pass image to itself and for matching the associated low-pass image to itself. The rationale for this procedure is that in recognizing the high-pass image from the low-pass (or vice versa), the available information is necessarily contained in the two images. By taking the average across the conditions in which these images had to be matched to themselves, one obtains some measure of the total information available to human observers trying to match them to one another.

Note that due to the compressive nature of accuracy, it is normally invalid to add or average such scores in this way. That is, generally speaking, it is more difficult to go from a score of 80 to 90% than it is to go from 60 to 70%. Because of this, accuracy cannot be considered an interval / ratio measure of performance, and it is therefore generally inaccurate to perform simple arithmetic on such scores in order to derive a measure of performance. However, due to the similarity of the findings for all filtering conditions in Experiment 3, only accuracies that were close in magnitude were averaged. This means that inaccuracies due to different levels of compression should be quite minimal.

An alternative possibility for deriving baseline values would have been to take the lowest of the two accuracy scores from Experiment 3 which corresponded to the previous experiments. That is, for each of the 9 conditions in Experiments 1 and 2, one could take the lowest associated accuracy score from Experiment 3 as the baseline. The rationale here is that this image provides the minimum amount of information available for recognizing the other image in the overlapping pair. This procedure would also avoid difficulties with compression, as discussed above. The rationale here is weaker, however, and the power of the design would be halved due to removal of data. Therefore, the baseline derivation method, as described above, was used in this study. An informal investigation of the data suggests that either method would arrive at qualitatively similar conclusions, however.

Overall, performance in Experiment 3 was quite good and fairly flat across spatial frequency bands, with even the most extreme filtering producing accuracies of over 75%. This suggests that the effects found in the previous experiments were indeed due to a lack of transfer between the learned and tested images, not the limited spatial range of the images themselves. That is, because the individual images could be matched to themselves with high accuracy, it could not have been the lack of information in the images themselves that caused difficulties in recognizing images with limited spatial frequency overlap.

Accuracy across conditions in Experiment 3 was also quite stable across conditions, producing a flat distribution in the baseline scores derived from them. The magnitude of effects of overlap seen in Experiments 1 and 2 can therefore be taken without modification due to the actual bandwidth or band location of the images in each condition. That is, there was little or no contribution of changes in the spatial bands of images per se to the effect of overlap.

Experiments 4 and 5 examined simultaneous matching of face images under filtering conditions similar to those in Experiments 1 to 3. Here subjects were shown two images of faces on the screen at the same time and asked if they represented the same individual. The only major difference between Experiments 1 to 3 and Experiments 4 to 5 (aside from the basic methodological change) was that both test and baseline conditions were included within subjects. That is, each subject was run through trials in which he or she was asked to match images with limited overlap as well as trials in which overlap was set at 100%. As expected, performance was much higher overall in Experiments 4 to 5 than in

Experiments 1 to 3 due to the lack of a memory load in the task and therefore the absence of mnemonic trace degradation and/or retrieval interference between stored representations.

Experiments 4a and 5a examined lower levels of overlap (from -2 to 0 octaves) and showed that only under the most extreme conditions (-2 octaves overlap at the 5.3 c/o center frequency) did matching performance approach chance. In general, performance rose smoothly as overlap increased. Experiments 4b and 5b examined higher levels of overlap, finding a continued trend towards increased performance, which generally reached ceiling at 2 octaves of overlap. Where performance reached ceiling, however, differences between conditions were still detectable in terms of reaction time and continued to show increased performance (decreased reaction time) as overlap increased. Accuracy in control conditions was at ceiling, which is to be expected from a simultaneous matching task of this nature, but reaction time data supported the findings of the previous experiments, showing little difference between conditions.

These findings suggest that overlap variations have approximately the same effects whether a subject is comparing two readily available images or comparing a single image with a stored representation. This indicates that the effects of overlap do not interact with mnemonic factors. There is no evidence that overlap has more or less effect based on the amount of time between the learning and testing stages. The fact that control conditions continued to show a fairly flat function of accuracy with spatial frequency band leads to the interesting suggestion that different bands of frequencies are retained in a similarly durable fashion and that no particular band is retained in preference to the others, as one might expect if a given band is innately more useful to the task of visual recognition.

Experiment 6 examined the effects of spatial frequency overlap on both face and object recognition in a sequential matching paradigm. Here subjects were shown two successive images of either a face or a chair, which were separated by a briefly-presented mask. Both test and baseline conditions were run. As expected, overall performance was between that of the previous two paradigms, with most accuracy levels around 70%. As with previous experiments, it was found that spatial frequency overlap had a strong effect on face recognition, but it had a non-significant effect on object recognition under the same circumstances. Indeed, switching from control conditions to experimental conditions changed performance in object recognition very little.

For both faces and objects, control conditions showed a small but steady increase in performance across spatial scale, but this was much low in magnitude per unit of change than the effect of overlap. This pattern of results is not surprising, considering that the amount of objectively available information in a given bandwidth is greater at higher frequencies than lower. Tendencies towards similar rises were seen in the previous experiments, but in all cases these were smaller in magnitude than the effects of overlap changes. These effects can therefore be assumed to have contributed little to the effects of overlap observed in the test conditions.

The baseline values derived from the control conditions show a flat function. This finding is not surprising, as the controls show a steady rise across spatial scale. Taking the average of two values at the extremes of this slope will necessarily give about the same value as taking the average of two values near the middle of it. Creating baselines in this way is valid only if one assumes that the difficulty in matching the two images is an average of the difficulty of matching each image to itself. Although this method of creating baselines seems plausible, another possibility would be to take the lowest of the two scores associated with a given overlap level, which, in this case is always that elicited by the lowest band. Following this method, the control conditions would show a slight increase, suggesting that spatial frequency overlap effects were not as large as indicated by the control conditions. The rise across controls is small enough, however, that significant rises in overlap would nonetheless be observed. In either case, the baseline data show that overlap effects are not due to a simple increase in the amount of information in the images, as increasing overlap is accompanied by both a rise and a drop in the spatial frequency range of the images to be matched.

Experiment 7 showed that results for upside-down faces were qualitatively similar to those for upright faces, with a steady rise in performance as overlap increased. This suggests that differences in how overlap affected faces and objects in Experiment 6 was not due to the qualitative nature of what was being shown in the image, but rather by stimulus characteristics that were unaffected by what the image represented. If faces were truly different from objects in terms of how they were affected by spatial frequency overlap, one would expect that upside-down faces would show a pattern different from that for upright faces. The reason for this prediction is that upside-down faces typically elicit non-face-like performance in other respects (Bartlett & Searcy, 1993; Farah, et al., 1995; Hole, George, & Dunsmore, 1999), and are thought to be treated more like objects by the visual system (Haxby et al., 1999; Aguirre et al., 1999). While the shape of the data for upright and inverted faces was similar, an overall drop in performance between Experiments 6 and 7 was as expected, matching many previous findings of extreme difficulty in recognizing upside-down faces (Bartlett & Searcy, 1993; Enns & Shore, 1997; Farah, et al., 1995; Parr, et al., 1998; Phelps & Roberts, 1994; Pullan & Rhodes, 1996; Tomonaga, 1994; Vermeire & Hamilton, 1998; Wright & Roberts, 1996; see Valentine, 1988 for a review of earlier work; Yin 1969, 1970).

Overall, the results show that, for face recognition, spatial frequency overlap has a strong effect. This is affected by position along the spatial scale, with higher frequencies producing higher accuracy. This is most likely due to the fact that a given band of overlapping frequencies at a higher point in the spectrum contains more information (Parish & Sperling, 1991). Conversely, changes in band location on the spectrum (when matching images with similar frequency ranges) had a weaker effect. In object recognition, on the other hand, the pattern of results is different. Here, the effects of overlap and frequency content were about the same. This argues that face recognition is somehow more vulnerable to changes in spatial frequency changes than is object recognition.

Comparing these results with studies that have examined similar questions in the past, one finds that there is good agreement between experiments. Millward and O'Toole (1986) found an accuracy level of 63% in a recognition paradigm using uncalibrated filtered face images in a condition where overlap was 0 octaves and the center frequency was 11 c/o. This was very similar to the 60% value found under similar circumstances (0 octaves overlap at 12 c/o) in Experiment 2a. Experiment 1a showed higher accuracy under the same circumstances, likely because the calibrated stimuli there did not contain aberrant frequency content that interfered with recognition. Also compatible with the results of the experiments presented here is Millward and O'Toole's (1986) finding that matching two similarly-filtered images to one another (i.e., low-pass to low-pass or high-pass to high-pass) was easier than matching either low-passed or high-passed images to a full-bandwidth image. As well, the recognition rates for low-pass to low-pass or high-pass to high-pass recognition were quite similar to one another, a result which is in line with our control conditions.

Biederman and Kalocsai (1997; Kalocsai & Biederman, 1998) found that face images were more vulnerable to changes in spatial frequency content than were chair images. In their equivalent of the control conditions tested here, the images at learn and test had a spatial frequency overlap of 100%. Under these conditions, both stimulus types were recognized very well, and elicited roughly equivalent performance relative to baseline conditions.. On the other hand, in their test condition where overlap was limited, they found that face recognition suffered greatly from lack of common spatial frequency content, with error rates of 15% compared to 8% in the 100% overlap condition (note that because exact figures were not provided by authors, estimates derived from figures are used in this discussion). This finding is compatible with the results of Experiment 6. However, their explanation for their findings is incompatible with the results of Experiment 7, which found little change in the effects of overlap when faces were inverted. This suggests that the difference between faces and objects was likely due to a greater degree of cross-scale correlation in object images than in chair images.

In summary, the findings of these experiments show that spatial frequency overlap has a large enough effect on recognition that it must explain a significant portion of the effects seen in other visual recognition phenomena. These include recognition across display format, coarse-quantization effects, and some differences between face and object recognition. The contribution of the overlap factor to these effects is examined in subsequent sections.

Recognizing images across display format

An understanding of spatial frequency overlap effects may help explain findings in the area of recognition across display format. Much previous research has found, for instance, that line drawings are adequate for object recognition but not for face recognition. Several explanations have been offered for this phenomenon. For instance, it has been argued that line drawings do not provide a proper impression of the mass of a face (Bruce, et al., 1992). Another explanation holds that line drawings reduce the ability for holistic configural processing to take place (Leder, 1996). Although these higher-level factors are likely part of the explanation, the findings of the present study argue that a significant portion of the effect is accounted for by the lower-level factor of spatial frequency overlap. That is, line drawings may simply not have enough spatial bandwidth in common with the internal representations of faces to which they are typically being compared. These internal representations are generally extracted from full-bandwidth images and therefore will likely contain elements from the full range of the spatial spectrum.

In many face perception studies, performance with line-drawing stimuli is compared to that with photographs. According to the results of the spatial frequency overlap studies in this dissertation, it may be that frequencies in the photographs that are not in the line drawings are interfering with recognition. That is, all other things being equal, a line drawing will have only a certain range of high frequencies in common with the stored image. The remaining lower-frequencies in the photographic image have no matches in the line drawing. Indeed, while these lower-frequency elements are unlikely to mask higher elements in a low-level fashion, they may interfere with recognition by adding noise to the process of higher-level feature matching. This is because of the high number of "misses" that will be scored in attempting to find matches in the line drawing for elements from the photograph.

It is important to note that the above explanation does not simply apply to direct comparison of two images. It also applies to the comparison of an input image to an internal representation. For instance, some studies comparing face recognition with line drawings and photographs have asked subjects to recognize famous faces. Subjects are not asked to learn the faces explicitly, as they are assumed to be familiar from popular media. As famous faces are generally learned from photographs or on television, the stored representation of the face should contain elements encoded from frequencies crossing the entire spatial spectrum. This results in a similar situation to that in which subjects view previously unknown faces in photographs and then attempt to recognize them in line drawings. That is, the internal representation in both cases is full bandwidth while the line-drawing is not. Similar limitations in overlap are found in both cases.

Other researchers have noted similar problems of comparability. For instance, Rhodes, Brennan and Carey (1987) point out this sort of problem in research on caricature recognition, noting that comparing performance on veridical photographic images to that with exaggerated line-drawings is problematic. This is because it is difficult in this case to know if effects seen are due to a switch from veridical to exaggerated images or to a switch from photographs to line-drawings. A better comparison, the researchers suggest, is between veridical and exaggerated line drawings.

Human observers' difficulties with forms of representation other than line drawings may also be explained in terms of spatial frequency overlap and interference from non-overlapping frequencies. For example, subjects typically have trouble recognizing bi-quantized faces, though this difficulty is not as pronounced as with line drawings. Here the internal representation is again (typically) full-bandwidth, but the image being presented contains only low frequency elements from the original. The high-frequency elements it does contain are not found in the internal representation. As with the line-drawing example given above, this means that there will be many misses when attempting match elements of the stored representation of a celebrity's face to the input provided by the bi-quantized face. This should explain at least part of the deficit seen when subjects attempt to recognize faces from two-tone images.

Given that this is the case, why is it that object images are recognized relatively well in these alternate forms of representation? Experiments 6 and 7 suggest that it may be due to object images having a greater degree of cross-scale correlation. In Experiment 6, faces showed vulnerability to spatial frequency variations while object recognition did not. In Experiment 7, however, inverted faces showed the same degree of vulnerability as upright ones. This suggests that the difference in performance between upright faces and objects is due to stimulus characteristics and not some difference in the way the face stimuli are treated. One likely difference between the two classes of stimuli is in terms of cross-scale correlation. Object images tend to have more sharp edges, features which transcend spatial scale. Thus, the identifying features of an object will typically be seen at a wide range of scales, whereas those that identify faces (shading and smooth shape variation) will be present across a smaller range. For this reason, if faces are presented in a mode that shares little spatial frequency range with the internal representation, there will be great difficulty in recognizing it. This is the situation that arises when trying to match a line drawing to a photograph, for instance. Conversely, line drawings of objects will have many elements that correlate strongly with the internal representation.

It is interesting to note in this context that coarse-quantized images may be viewed as simply another form of representation. Like bi-quantized images, they remove the high spatial frequencies from the original while adding new high-frequency structure. Unlike the bi-quantized image, however, there is little correlation between the newly added elements and the original elements that have been removed. This means that not only does the new image lack elements that overlap with those in the original, but also it has a set of features (the square grid) that find no match in the original. This creates excessive difficulties in recognition unless the grid pattern is either minimized in saliency or is somehow segregated from the face pattern.

In studying this sort of image. Harmon and Julesz (Harmon, 1971; Harmon & Julesz, 1973) argued that the effect of coarse-quantization was due to added high-frequency noise. In contrast, Morrone and colleagues (Morrone et al., 1983; Morrone & Burr, 1994) argue for a more high-level explanation involving the misplacement of features in the image. The results of the present experiments suggest that a basic contributing factor to the difficulty in recognizing coarse-quantized images is the reduced correlation between the internal representation of an image and the input that results from this treatment of the image. Not only are the original high frequencies taken out of the image, but new ones are added in, causing the original image and the pixelated one to have very different energy distributions.

While spatial frequency overlap effects may help explain difficulties in matching across various display formats, this explanation is certainly not the whole answer. Different forms of representation are most likely innately superior in the amount of information they provide to the visual system. One may find, for instance, that although drawing-to-drawing matches are better than photograph-to-drawing ones, matching one photograph to another is still superior. In general it is not surprising that matching between two images in the same display format will be easier than matching between images in different display formats. But the important point here is that the degree of difficulty posed by cross-representation recognition will largely determined by the degree to which the images overlap in spatial frequency content. Matching a low-passed or coarse-quantized image to a line-drawing, for instance, may prove extremely difficult, due to the fact that these two formats will not share many informative elements. This effect is seen both when doing simultaneous matching and recognition of a learned picture from a test picture, demonstrating that the effect is present both with regards to image inputs as well as cognitive representations.

A possible limitation of Experiments 1 to 5 is that the pose of the face did not change between learning and testing. Because of this, one might argue that what was examined was the ability of subjects to perform general image-based recognition tasks, as opposed to face recognition per se. In their discussion of the difference between image-recognition and item recognition, Bruce and Young (1986) suggest that the latter depends on two codes: A pictorial code that is relatively concrete and image-based, and a structural code that is more abstract and depth-based. The pictorial code includes situation-specific elements arising from the effects of lighting, pose, and so on, while the structural code incorporates the invariant elements that identify the face under various viewing conditions. Based on these descriptions, Experiments 1 to 5 could be said to have examined effects within the domain of the pictorial code rather than the structural and therefore may not be relevant to real-world face recognition. Although similar results were obtained in Experiments 6 and 7--where faces were rotated 45 degrees between learning and testing--assuage this concern to a large degree, a discussion of this potential limitation nonetheless seems mandated.

In Experiments 1 to 5, the pictures at learning and test were identical except for the way in which they were spatially filtered. Thus, the same lighting and pose factors were present in both cases. This means that there was little need for the structural code to be accessed in order for correct recognition or matching to occur. It also means that spurious elements of lighting and so on could be used to aid identification. Because the structural code is an invariant identifier of individual objects, it may be argued that this aspect of visual object representations is more relevant to the question of how visual recognition takes place, whereas the pictorial code may be thought of as containing information that is irrelevant to the real-world execution of this problem. That is, because the pictorial code can contain transient information that is not an intrinsic part of the object being represented, it may incorporate a certain degree of "noise" for the task of invariant face or object recognition.

Contrary to this line of argumentation, a number of studies have used methodologies that emphasize the use of the pictorial code in recognition and have found face-specific or object-specific effects. For instance, Biederman and Kalocsai (1997) used tasks similar to those used in Experiments 1 to 5 of the present study, but found differences in the behaviour of face and object stimuli. In their study, face recognition was more affected than object recognition by the absence of overlapping frequencies between the learn and test images. This shows that image-matching is involved in higher-level processes in visual recognition and is affected by what the image represents.

Tarr, Kersten and Bulthoff (1998) suggest that lighting, a factor thought to be part of the pictorial code, may in fact be encoded as part of an object's invariant representation. They find that lighting 3D objects from different sides affects recognition in a way that suggests lighting effects are stored in visual memory and that this is done in order to disambiguate the shape of objects. They further suggest that lighting-based features are not only stored in terms of their effects on the image being input, but actually in terms of the object's invariant 3D representation. If this is the case, it suggests that the pictorial code is highly relevant even to recognition even under varying viewing conditions.

Although the above arguments support the importance of pictorial cues in real-world recognition, one cannot deny that the structural code will often be more important to the task. But the fact that pictorially-based tasks show differences based on the content of images, and the fact that certain pictorial cues seem to be stored in object-memory, suggest that any differences between the relevance of the two codes will be relative. Both will play a part in most recognition tasks to varying degrees. Accepting this, the interesting question then concerns under which circumstances one or the other plays a more vital part.

One factor that may determine the relative usefulness of pictorial vs. structural codes is familiarity with the faces being recognized. A number of studies have found that for unfamiliar faces, changes in elements of the pictorial code (such as lighting direction, pose and so on) are detrimental, but with familiar faces, these have little effect (Kemp, Towell, &

Pike, 1997; Bruce, et al., in press; Burton, et al., 1999). Presumably this is because structural codes have been obtained in the case of familiar faces and provide relative robustness to variation. This suggests that recognition of unfamiliar faces is governed primarily by pictorial codes, making an understanding of this form of recognition very important for understanding recognition of faces not previously seen. Based on this, the data from Experiments 1 to 5 of the current study can be seen as applying primarily to an understanding of how one acquires representations of new faces, while Experiments 6 and 7 may have broader applicability.

Experiments 6 and 7 examined conditions under which structural codes might be expected to be more important than pictorial ones. That is, these experiments examined sequential matching under conditions where the images at learning and test were rotated 45 degrees relative to one another. Pictorial cues based on pose were therefore not useful for recognition and the task depended more on invariant properties of the faces. Under these conditions, the effects of overlap were quite similar to those seen when the images were in the same pose at learning and test (i.e., in Experiments 1 to 5). This lack of an interaction between the overlap factor and the use of pictorial vs. structural codes assuages concerns that the effects seen in the earlier experiments might have been due to transient factors involving lighting and other image properties. It also suggests that similarity in spatial band between learning and test is an important factor even when higher-order functions such as rotation in depth are required to perform a perceptual task. Thus, knowing about how differences in signal input affect recognition is also important in cases where the structural code is involved. One factor that determines such similarities is spatial frequency overlap.

In summary, it is clear that studies that examine how mode of representation affects recognition can benefit from an account based on signal similarity between the learned and tested image. One aspect of this similarity was examined in this dissertation: Range of shared spatial frequencies, or spatial frequency overlap. The effect of this factor is strong enough that it may account for a significant portion of the effects seen when going from one display format to another. This was the case with faces more than with objects. Face effects were seen whether images were pairs of band-passed photos, or high-pass vs. low-pass pairs. Effects also remained stable across a variety of methodological differences.

Which bands of spatial frequencies are most useful?

Early interest in the effects of spatial filtering on face and object perception focussed on the question of whether high or low frequencies were most important for recognition. Ginsburg (1978, 1980) argued that low frequency elements were of primary import in recognition, with higher frequencies being redundant. Later researchers, such as Fiorentini et al. (1983) argued that higher frequencies were sufficient for recognition and were not redundant to the process of recognition. Recent studies suggest that for face recognition in particular, a middle band of frequencies between 8 and 16 c/o are of primary importance to that perceptual task (Bachmann, 1991; Costen, et al., 1994, 1996; Gold, et al., 1999; Nasanen, 1999).

In contrast to these studies, the results of Experiments 3 to 7 do not support the notion that any particular band of frequencies is most useful for face or object recognition. The control conditions tested in these experiments are most relevant here, as subjects were required to match images that had been similarly filtered to each other. These conditions all showed a flat or slightly rising function of performance with changes in spatial frequency content both within and outside the critical bands suggested by previous research. This was the case for low-passed, high-passed and band-passed images. Also, the pattern of results remained the same whether stimuli were objects or faces, whether they were rotated in depth or not, and whether the paradigm used was simultaneous matching, sequential matching or recognition. The slight rise in performance with increasing spatial frequency might be seen as support for experiments that have found higher spatial frequencies to be of greater importance for recognition (e.g., Parker et al., 1996), but it is difficult to make this conclusion, as the objectively available amount of information in a given spectral range rises with spatial frequency.

The control conditions of Experiments 3 to 5 used low-passed and high-passed stimuli. Subjects were required to recognize or match pairs of images that had been filtered at a series of different cut-off points. An interesting finding is that performance did not show a rise and fall pattern as the spatial range of images expanded into and beyond the critical band for face recognition, as would be expected based on the "critical band" studies. Rather, performance was quite flat, with only a slight and inconsistent rise as the bands expanded. Experiments 6 and 7 showed that when subjects matched images that were filtered to contain the same limited band of frequencies, there was no particular advantage for bands with center frequencies in the critical region. As with previous experiments, there was only a moderate increase in performance as the frequency of the band rose.

The most probable reason for the differences between the findings of the present studies and those of past reports seems to be that here subjects were being asked to match images that had been filtered in the same way, whereas past experiments asked participants to match filtered images to full-bandwidth ones. The present findings suggest that the critical bands of frequencies found by previous researchers are only of import in a context that includes the rest of the frequencies in a full-bandwidth image. That is, there may be something about these frequencies that allows them to be extracted from a full-bandwidth image more easily, or they may somehow be interfered with less by other frequencies. As was suggested earlier, one problem with recognizing a spatially-filtered image from a full-bandwidth image is that the latter contains a wide range of frequency elements not found in the former. It is possible that the middle frequencies are in some way less affected by this sort of high-level interference.

Previous explanations have suggested that middle frequencies are most important because of the structure of the face (Costen et al., 1994, 1996). But if this were the case we would expect that matching pairs of faces band-passed to contain mostly frequencies within the critical would be easier than matching those with only frequencies below or above. This is not observed in these experiments.

To this point, the relevance of these findings to recognition across display format has been examined from a highly theoretical viewpoint. But it is perhaps more relevant to look at the practical implications, particularly as this has been the focus of much of the research in this area.

There are a number of practical consequences that arise from the results of this study. For instance, the present findings argue that if one wishes to have people recognize a face image, it is important for the input and the representation to not only share spatial bandwidth, but also to avoid having excessive spatial frequency information outside each other's ranges. As an example, researchers have shown that matching a full bandwidth photograph to a low-quality video camera image is extremely difficult, even for forensically trained police officers (Burton et al., 1999). Based on the findings of the present study, subjects should paradoxically be better able to recognize the video camera image from a similarly degraded stock photo.

Of course, it would not be surprising or of practical interest if degradation merely rendered a set of images more homogeneous in a higher-level sense. That is, such degradation of images might simply render the set of images indistinguishable from one another. If this were the case, one would expect to see a pattern of results wherein both hits and false alarms vary in concert with the degree of degradation of the image set. But according to the interpretation offered here, one should instead see an increase in overall measures of accuracy, with a higher number of both hits and correct rejections as the images come to have more of the spatial frequency spectrum in common. The reason for this is that such a manipulation would increase the amount of common information available for making a match, and decrease the amount of information in bands that the images did not share.

The findings of the present studies also have a number of implications for experiments that draw conclusions about spatial frequency content from tasks utilizing recognition across representational form. For instance, Biederman (1987) has suggested that high-frequency information may be sufficient for object recognition, based on studies of line drawings. But, in the current study, no evidence of a special role for these frequencies was found. In experiments with both face and object stimuli it was found that a wide range of frequency bands were sufficient for recognition, and that any differences based on spatial frequency changes were similar in character for both faces and objects (i.e., in terms of band of frequencies occupied, there were marked differences in terms of spatial frequency overlap effects).

Some researchers have suggested that line drawings are adequate for object recognition but not for face recognition (e.g., compare Bruce, et al., 1992 and Davies, et al., 1978 to Biederman & Ju, 1988). They suggest that this is due to the importance of low frequencies for recognition of the latter. But again, the data of the present experiments does not support this. Rather, they support the notion that line drawings are adequate for object recognition because of greater cross-scale correlation in the sorts of stimuli used in most object-recognition experiments.

It is clear from the above that knowing about which bands of frequencies are best for allowing recognition of unfiltered images from filtered ones does not give us a complete understanding of the visual recognition process. Humans are sometimes faced with situations where two images are bandwidth limited, and yet they are capable of matching between them. An interesting example of this can be seen by considering the Burton et al. (1999) study, as described earlier, but imagining that the subjects involved were required to match sketches of individuals to security camera photos. The former image will tend to present relevant information at only high frequencies, whereas the latter will generally contain relevant information at only low frequencies. Both images might contain frequencies in the critical band, but they would still be extremely difficult to match with one another due to the lack of frequencies they share.

In summary, it seems clear that an understanding of how similarity in spatial frequency range affects recognition can be informative for understanding the effects of varying mode of representation. The effects of spatial frequency overlap on recognition and matching performance are sufficient to account for a significant portion of the differences arising from changes in display format. This factor, along with other signal quality differences, must be taken into account in order for a complete understanding of representation effect to be had. Spatial frequency overlap also has relevance for some practical problems, such as those encountered in security situations where individuals must match highly different representations of an individual.

How does luminance calibration of filtered stimuli affect performance?

Several authors have noted the importance of presenting spatially filtered stimuli on a calibrated monitor (Metha, et al., 1993; Olds, Cowin & Jolicoeur, 1999; Peli, 1992a; Pelli & Zhang, 1991; Tyler & McBride, 1997). The reason for this is that the majority of monitors, in their native state, exhibit a relationship between luminance and voltage level that is non-linear. Generally the lower voltage levels (or grey-scale levels) represent small increases in luminance, whereas the higher ones represent large increases. When dealing with spatial filtering, this non-linearity results in filtered images having additional energy added to them. This is especially the case with high-passed images where a great deal of energy can be added to the image at the low end of the spectrum. Surprisingly, although there have been a number of articles on how to calibrate monitors and why it is recommended, there have been no studies to date that have explored the actual differences calibration might make in human performance when recognizing complex images.

In order to examine the effects of luminance calibration as they interact with spatial frequency overlap, Experiments 1 and 4 were repeated with non-calibrated stimuli in Experiments 2 and 5, respectively. Generally the results were in good agreement, no statistically significant interactions were seen between the calibration factor and others. The exception to this was that Experiment 1a (calibrated) showed easier recognition in the 0 octaves of overlap condition as compared to Experiment 2a (uncalibrated). Here a clear effect is seen whereby all 0 octave of overlap conditions showed above-chance performance in the calibrated version, but only the 0 octave of overlap condition at center frequency 12.0 c/o showed above-chance performance in the uncalibrated condition. The difference in accuracy was as high as 9%. As all the studies show an increase in performance with increasing center frequency, it can be reasoned that there was an overall suppressive effect on recognition in the uncalibrated case. But why did this only occur for the recognition conditions and why only with low levels of overlap? One possibility is that the difference arose because of the different challenges posed by the tasks of matching across a gap in spatial frequency as opposed to an overlap in frequency. In the former case, a matching algorithm cannot look for similar elements in the two images because there are none. Even if one postulates a simple template-matching algorithm with perfect correction for spatial location, rotation, and so on, an algorithm that only looked for elements of similar spatial frequency (i.e., detected by the same early channel) in the two images would be left with no hits and be completely unable to make matches. Instead, a system must be hypothesized that takes advantage of the correlation of information across spatial scale and that allows for hits be scored when image elements have the same relative spatial location and similar orientation, regardless of moderate differences in spatial scale. Unfortunately, such an algorithm would be vulnerable to the inclusion of noise, especially the kind of noise introduced by non-calibration. The reason for this is that the algorithm would allow incorrect matches between the aberrant spatial elements in the image and similar elements in the representation, resulting in incorrect confirmations (false alarms).

Conversely, in the case of an overlap in spatial frequency, the algorithm can be much more conservative in accepting elements as matching. This is because there are in fact elements that are similar in spatial frequency between the two images. These can be used to assess degree of match without the need for correlating elements across scale and therefore without vulnerability to the noise elements introduced by non-calibration of stimuli. In short, when there is a gap in spatial frequency content between two images, only correlative information is available and matching procedures must be fairly lax in making matches between features, making them vulnerable to noise. On the other hand, when there is an overlap in spatial frequency the matching procedure can be more conservative and thus is protected against noise. Note that the simple template-matching algorithm presented here is considered for the purposes of example only and is not necessarily a good model of recognition, but similar basic principles should apply to any matching algorithm.

Many early studies have overlooked the issue of screen calibration. Most of the studies discussed in the literature review do not mention the issue at all, leading one to suspect that at least among early studies the stimuli were presented on monitors using only the native grey scale values. Modern studies are typically assumed to use calibrated stimuli, but again most authors do not make this explicit. This is especially the case with studies that use complex images. Therefore, it is likely that the literature contains a mixture of studies variously using calibrated and uncalibrated presentation methods. Especially problematic is a possible division between those studying lower-level processes, who typically acknowledge the necessity of calibration, and those studying higher-level processes, who may consider calibration a minor technical issue with no effect on complex higher-level tasks. If such a divergence exists, it could create difficulties in using studies in one area to explain results in the other. While the data in the present study suggest that effects will be fairly minor under most circumstances, it is nonetheless important that low-level and high-level studies be methodologically comparable if they are to gain insight from one another.

There exist to date no other studies that have investigated the question of the effects of screen calibration on recognition of complex filtered images. For this reason, there can be no comparison with previous results. The present set of experiments suggests that studies that have used calibrated images and those that have not may be compared with moderate confidence, although small reductions in accuracy may be expected for non-calibrated stimuli. It is important to note that these assertions regarding the effects of calibration on recognition can only be made with regards to face perception, as that is what was tested here. The effects on object recognition were not tested, although one might predict that they would be lesser for images having greater correlation across scale.

In summary, calibrating one's stimuli seems to have little effect on complex image recognition except under particular circumstances. This is likely due to redundancies in natural images that the visual system is able to take advantage of in most tasks. Only where the degree of redundancy is attenuated will calibration have an effect. Few studies have investigated the effects of screen calibration on complex image recognition, and many studies in this area fail to acknowledge the importance of using calibrated stimuli. The data here suggest that this will cause only minor and limited discordance in results, but effects are sufficiently large to mandate the use of calibration even in studies of higher level processes.

How does task type affect recognition?

Early research examining whether high or low frequency information was of greatest import to visual recognition is contradictory. Sergent (1986) attempted to reconcile these findings by appealing to their different methodologies. She noted that while Ginsburg (1978; 1980) found support for the importance of low frequencies using a recognition paradigm, Fiorentini et al. (1983) found support for higher frequencies using a matching paradigm. She suggested that the former task relied more on the overall shape of the face or object being recognized, whereas the latter relied more on the interior features and finer details of the face. Thus, Ginsburg's (1978; 1980) methodology emphasized low-frequency elements in the image while Fiorentini et al's (1983) emphasized high-frequency ones. Although this explanation is appealing, it does not bring order to the results in the literature.

In the present study, three paradigms were used: Recognition, simultaneous matching and sequential matching. All of these paradigms showed roughly similar effects of overlap, though overall performance dropped as latency between learn and test images increased. This argues against the notion that task-type might affect which spatial frequencies are most useful in recognition. In these experiments we found only that higher frequencies were more useful for recognition than low. Performance steadily increases as the overlapping spatial frequencies in the images to be matched become higher. Also, with both face and object matching there is a rise in performance as the band mutually occupied by images for matching increases.

This pattern of results argues against the idea that face matching is subserved by the external features of the face while face recognition relies on internal features. Rather, the data presented here argue that in all forms of face perception, different spatial frequency bands serve with roughly the same efficiency. How then can we explain the apparently contradictory findings in the literature? It seems from the present findings, that a wide variety of spatial frequency bands are sufficient for recognition and that no given band is necessary for recognition. What is necessary for recognition is a sufficient degree of overlap in spatial frequency range. For low-pass and high-pass images being matched to full-bandwidth pictures, this amounts to saying that a sufficient bandwidth in the filtered images is required. For band-passed images, it amounts to saying that the bands they occupy must overlap to a sufficient degree.

This is the same whether one is attempting to match a stored image to an input one or one is attempting to match two input images simultaneously. Note that in both of these cases there is in fact a stored image which is being compared to the input image, it is only the latency between storage and comparison that is being changed. The data from the present study suggest that no matter what the latency might be, a fundamentally similar process takes place. The comparison between the two images does not differ depending on whether one image is in short term storage and the other in long term or both are in short-term. A condition where both images are in long-term store was not tested, but it may be predicted that similar results would be obtained.

The only pattern in the current data that points to a difference based on task type is a possible interaction with calibration effects. In recognition conditions (Experiments 1 and 2) calibration had an effect, with calibrated stimuli producing superior results to uncalibrated in low overlap conditions. However, in simultaneous matching conditions (Experiments 4 and 5) there was no such difference. This may point to different processes underlying the two forms of matching, but more likely this finding reflects the fact that the effects of calibration are small and were likely too insignificant to come out in the easier matching conditions.

In summary, the data found in these experiments do not support the idea that task type affects the spatial frequency range that is important to face recognition. In all cases, it was found that a wide range of spatial frequency bands are sufficient for face recognition and that none are necessary or particularly advantageous.

Do objects and faces differ in retention of spatial frequency information?

Biederman and Kalocsai (1997; Kalocsai & Biederman, 1998) propose a model of face and object representation in which faces retain spatial frequency information while objects do not. This, they argue, is the reason for the relative robustness of object recognition to variations and degradations in spatial frequency content. In their model, faces are stored in terms of metric variations in a vector of low-level filter activation levels. To illustrate Biederman and Kalocsai's model, consider a simple model of the visual system as an array of wavelet detector "pyramids". The lower levels of each pyramid contain many detectors with small spatial extents whose purpose it is to detect high-frequency elements in the retinal image. As one moves up the pyramid, each layer contains fewer detectors, but with larger spatial extents, whose purpose it is to detect low-frequency elements. At each point in a given level, a number of different orientations of detector are present.

When an array of such detectors is presented with a simple image of an edge feature, only some of them at a given spatial location will activate. Activation will be limited to those detectors whose receptive fields are collocated with the edge in the retinal image. It will also be limited to those that are roughly of the same orientation as the feature. If the edge is sharp, the activated detectors will cross a wide range of spatial frequencies. If the edge is blurred, only the low frequency detectors will be strongly activated.

A face image presented to such an array produces a complex pattern of activation, which can be stored as a vector of relative values. Each value represents the activation level of a certain feature detector. Given that factors such as pose and position in the scene are the same, each individual face will create a roughly similar vector of values, with little in the way of qualitative differences between them. For this reason, Biederman (1987; Biederman & Kalocsai, 1997; Kalocsai & Biederman, 1998) holds that individual face representations must consist of variations from a metric (i.e., differences in vector values from the vector values of a "standard face"). This sort of representation, because it stores data about the activation level of every detector in the array, retains information about spatial frequency content in the image.

In contrast to faces, Biederman (1987; Biederman & Kalocsai, 1997; Kalocsai & Biederman, 1998) holds that the common objects that humans can quickly and easily identify as distinct tend to have broad qualitative differences. That is, objects are held to differ from one another in terms of arrangements of "non-accidental properties" or NAPs. Examples of NAPs include pairs of parallel lines--which will tend to stay parallel across a wide variety of viewing angles--and curves--whose convexity or concavity will likewise be maintained over a range of perspectives. NAPs are the basis used to derive Geons, object components consisting of volumetric primitives such as bricks, cones and cylinders. Object representations consist of lists of such primitives with relational statements linking them. For instance, a typical hand-held camera might be described by two Geons, a cylinder and brick, with a single relational statement--"In front of"--linking them (i.e., the cylinder is in front of the brick). Most objects can be described by two or three geons, according to Biederman (1987). An object representation is known as a Geon Structural Description, or GSD.

Because NAPs and GSDs draw their initial data from object edges primarily, it is not as important for object representations to maintain spatial frequency information. In fact, Biederman holds that in forming object representations the brain discards spatial frequency information. Only the spatial location of a feature is retained. In the model of vision presented above, only the fact that a given relative spatial location has had a detection is retained, not which layer or layers of the pyramid did the detecting.

Based on the above description, faces should show vulnerability to reductions in spatial frequency overlap, whereas objects should not. The reason for this is that face representations retain spatial frequency. If an input face image results in the activation of a detector at the same spatial location as was recorded for a previous image, but not at the same spatial frequency, no match should be recorded (or more likely a "weaker" match would be recorded). For an object, however, the spatial frequency of the detector is not recorded, so a feature detected at the same spatial location and orientation should produce a match with a similarly located feature in the representation, regardless of the spatial frequency of the input and representational features. Of course, some features are eliminated in any filtering operation, but the edges that are supposed to be important to object recognition will generally remain despite filtering. Sharp edges tend to transcend spatial scale and are present in low-passed or high-passed images in nearly the same location as in the unfiltered version. Based on this, one would expect that as two band-passed images become more similar in the spatial range they occupy, faces should show a rapid rise in recognition rates, while object recognition will remain relatively stable.

This suggestion is indeed supported by the results of Experiment 6 of the current study. However, an alternate explanation for this pattern of results may be proposed. Specifically, it is possible that some stimulus-based aspect of face images vs. object images results in the differences. One likely possibility is that face images have a lower correlation of visual information across spatial scale. That is, if one filters a face image to contain only a given band of spatial frequencies and cross-correlates it with the same face image containing only a different band of frequencies, the correlation may be quite weak. On the other hand, it might be quite strong if a similar procedure were applied to object images. This seems likely to be the case, because object images tend to have more sharp edges than do faces.

If this is the explanation for the difference, then we should see equal vulnerability to spatial frequency overlap for faces when they are turned upside-down. If, on the other hand, inverted faces are stored in a GSD, we should expect low but stable performance across levels of spatial frequency overlap. The reason for this prediction is that upside-down faces, if treated as objects, would be expected to be represented as GSDs. These GSDs would be quite homogenous and therefore difficult to distinguish (thus low overall performance) but would be as unaffected by spatial frequency overlap limitations as other objects (thus little effect of this factor).

The results of Experiment 7 do not provide any support for this notion.. Instead, inverted faces showed nearly the same magnitude of changes due to spatial frequency overlap as did upright faces. Only an overall drop in performance was seen. This argues for an explanation based on stimulus characteristic differences. The most likely candidate for the relevant difference is degree of correlation across scale.

In summary, the results of Experiments 6 and 7 do not support Biederman and Kalocsai's (1997) model of face and object representation. Instead, the results are more compatible with an explanation in which both faces and objects are stored in a manner that retains spatial frequency information. This has important implications for the question of face "specialness", as it suggests that faces are not stored in a different form than objects. It should be noted that this only applies to this particular aspect of face specialness and does not address the question of whether faces are processed by a separate module in the brain. A great deal of evidence still suggests that faces elicit qualitatively different recognition behaviour from subjects, and these experiments do nothing to contradict this. The results reported here only suggest that faces and objects are stored in similar ways in terms of spatial frequency content, and say little about the possibility of higher-level perceptual or cognitive differences in how these classes of stimuli are treated.

Future Directions

A number of directions for future study are suggested by the findings of this study, some theoretical in nature and others practical. On the practical side, it would be interesting to determine if similarly degrading images would render them more recognizable from one another. Ideally, such a study would be very naturalistic in nature, involving an analysis of images from an actual security video camera as compared to a high-quality camera to determine the exact nature of the degradation involved. Commercial security devices appear to impose a number of limitations on image quality, including pixelization, defocus, reduction of contrast, glare effects, and motion blur, so this may be technically challenging. However, it should be possible to show that as one makes two images more similar in the manner in which they are degraded they become more easily matched from among a sample of similarly degraded images. As stated before, it would be important to show that degradation was not simply making images less distinguishable, but that it was genuinely increasing performance. Also, it would be important to determine how such effects interacted with those of changing the pose of the face between images, as security cameras tend to take their pictures from unusual view-points. This line of research has the potential to contribute to police and security work.

On a more theoretical level, the results of this experiment have suggested that there is no particular band of frequencies that are more useful than others. But this was found in the case of matching a filtered image to a similarly filtered image, as opposed to matching a full-bandwidth image to a filtered one. It would be interesting to do a direct comparison of these two sorts of conditions to determine the nature of the apparent higher efficiency of certain bands. Such work would more directly test the assertion that the relative import of these bands was due to interactions with the rest of the spectrum included in a full-bandwidth image.

To more exhaustively examine this question, one could also study the effects of varying the range of extraneous spatial frequencies on image recognition. For instance, one could have subjects match band-passed images to other band-passed images with the same center-frequency but different band-widths and measure how recognition degrades as these additional frequencies are added in. If the degree of degradation with frequency addition was great enough, one could conclude that much of the effect seen in the critical band studies were due to interference from extraneous frequencies in the full-bandwidth images. This would give us a more complete understanding of how spatial frequency factors affect recognition.

Although a further understanding of the sensory aspects of visual recognition is of great interest, the findings here also point to a number of more cognitively-oriented experiments that might give us a better understanding of visual memory. For instance, it has been suggested by a number of researchers that certain bands of spatial frequencies are most useful for certain tasks, such as recognizing faces or objects. If this is the case, we might expect that these bands would be retained better by visual memory than other bands. For instance, if the band of 8 to 16 cycles per face is best retained, then we might predict that the information at this scale would not be allowed to degrade as easily as information at other bands. The reason for this prediction is that, assuming there are limited storage resources and that degradation of is a means of economizing these resources, it would be most efficient for visual memory to discard less useful information. Of course, there are a number of assumptions underlying this formulation, one of which is that different scales are stored in such a way that they may degrade independently and that the visual cognitive system has a way of choosing which bands are most informative.

The results of this study do not support the notion of a given band being superior to another for face or object recognition, finding a flat function for recognition of band-passed images as the band occupied by the images shifts across spatial scale. This is the case whether images are being simultaneously matched or recognized from learned images, suggesting that no particular band is being retained better than others. But these findings do not arise from a direct test of the hypothesis and therefore cannot constitute a direct rebuttal of it. In order to execute such a test, one would have to have subjects learn a set of faces and then test them at several intervals afterwards under similar circumstances. The expectation, if certain bands are more efficient, would be that these bands would show a degradation curve that was shallower than the others. In line with the previously suggested experiments, it would be interesting to examine effects under two conditions: One where full-bandwidth images are recognized from filtered version, and another where two similarly-filtered images are recognized from one another.

Another possible way to expand on the present studies would be to gather data from an ideal observer set up to do the tasks and compare this to human data, thus obtaining efficiency scores for the various conditions rather than simple accuracy measures. This has the advantage of accounting for all the information in the images themselves and thus allowing firmer conclusions as to the source of effects (i.e., from the visual system or from stimulus qualities). Control of stimulus quality effects is a difficult issue in complex image recognition because the stimuli are, by definition, elaborate. Many studies simply ignore this issue, creating results that are difficult to interpret. In the present studies, extensive use of control conditions was implemented on the assumption that differences between these and the test conditions could be safely attributed to visual system effects due to the fact that the stimulus characteristics were the same in both cases.

This approach has both advantages and disadvantages relative to using an ideal observer as a baseline. With ideal observers, one gets a baseline which controls strongly for any information inherent in the image itself and subject accuracy is compared to this to obtain efficiency scores. Thus, any variations in efficiency must be due to factors within the observer. However, with complex stimuli it is difficult to narrow down what aspects of the observer account for the effects. They could arise at a number of levels, including cognitive, attention-based or low-level sensory ones. The difficulty arises from the fact that human observers bring with them a great deal of information about what objects "should look like." In the case of faces, for instance, they know that the stimuli should be symmetrical and that they should have a similar basic arrangement of features. They may also be able to significantly extrapolate on the information presented to them in the images. A basic ideal observer does not do this, and while it may be possible to build in such constraints and abilities, it is difficult to do so without first understanding how human observers recognize objects. Thus, the problem becomes circular. It is difficult to construct an ideal observer that is qualitatively similar enough to a human observer to help understand human observers unless we first understand human observers.

Having said this, however, there are also disadvantages to using human observers as baselines. These have to do with the fact that the task cannot be exactly the same in the baseline and experimental conditions. Because both methods have advantages, it would be most informative to have baselines from both. Ideal observer data would provide a stronger control for the amount of data in the images, and would be especially useful in comparing data across stimulus types (i.e., faces vs. objects) as it would allow a factoring out of the potentially large overall differences between these classes in terms of image qualities.

With regards to object recognition specifically, one direction that might be taken from the present experiment is to examine intra- and inter-class object recognition and the effects spatial frequency overlap has on these two sorts of visual discrimination tasks. While the distinction between inter- and intra- class recognition is widely recognized as important, there has been no work directly comparing how spatial frequency range affects these. The results of this experiment showed little effect of spatial frequency variations with object stimuli, but a great deal with face stimuli. If, as Gauthier (1999) has argued, face specific effects are contributed to by the greater homogeneity of this stimulus class as compared to the objects that have been used as controls for it, then we might expect that inter-class recognition would be even less affected by spatial frequency overlap.

Conclusions

This dissertation examined a novel aspect of spatial frequency effects on visual recognition, specifically spatial frequency overlap. Its results dealt primarily with face recognition, although object recognition was examined in two experiments. The major conclusion that can be drawn from this research is that spatial frequency overlap has a strong effect on recognition, and that this effect is greater than that of location on the spatial scale. Related to this is the conclusion that overlap has stronger effects on face recognition than on object recognition, most likely because the latter generally involves images with greater correlation across scale. A second important conclusion that may be drawn from this research is that the effects of luminance calibration on recognition accuracy will generally be small, except in cases where images sharing no spatial range are to be recognized from one another.

In conclusion, the data gathered here on the effects of spatial frequency overlap serve to expand knowledge in general about how we recognize the things we see. More specifically they enable us to understand more clearly how different forms of representation affect visual recognition. The data help to both clarify previous theoretical findings and to suggest directions for future practical research.

BIBLIOGRAPHY

Aguirre, G.K., Singh, R., & D'Esposito, M. (1999). Stimulus inversion and the responses of face and object-sensitive cortical areas. Neuroreport, 10, 189-194.

Anderson, J.R. (1994). Valeur ethologique des visages et des mimiques chez les primates non-humains. Psychologie Francaise, 39, 345-355.

Anstis, S. (1992). Visual adaptation to a negative, brightness-reversed world: Some preliminary observations. In G. A. Carpenter, S. Grossberg (Eds.), Neural Networks for Vision and Image Processing (pp. 1-14). London, England: The MIT Press.

Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? European Journal of Cognitive Psychology, 3, 85-103.

Bartlett, J.C., & Searcy, J. (1993). Inversion and configuration of faces. Cognitive Psychology, 25, 281-316.

Biederman, I. (1987). Recognition by components: A theory of human image understanding. Psychological Review, 94, 115-147.

Biederman, I. (1993). Geon theory as an account of shape recognition in mind and brain. Irish Journal of Psychology, 14, 314-327.

Biederman, I. & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. Philosophical Transactions of the Royal Society: Biological Sciences, 352, 1203-1219.

Biederman, I., & Ju G. (1988). Surface vs. edge-based determinants of visual recognition. Cognitive Psychology, 20, 38-64.

Blakemore, C. & Campbell, F.W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. Journal of Physiology, 203, 237-260.

Bradshaw, J.L., & Wallace, G. (1971). Models for the processing and identification of faces. Perception & Psychophysics, 9, 443-448.

Braje, W.L., Kersten, D., Tarr, M.J., & Troje, N.F. (1998). Illumination effect in face recognition. Psychobiology, 26, 371-380.

Bruce, V. (1988). Recognizing Faces. Hove: Lawrence Erlbaum Associates, Ltd.

Bruce, V. (1998). Fleeting images of shade: Identifying people caught on video. Psychologist, 11, 331-337.

Bruce, V., Burton, A.M., & Craw, I. (1992). Modelling face recognition. Philosophical Transactions of the Royal Society of London: B, 335, 121-128.

Bruce, V., Cowey, A., Ellis, A.W., & Perrett, D.I. (Eds.). (1992). Processing the Facial Image. Oxford: Oxford University Press.

Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The importance of 'mass' in line drawings of faces. Applied Cognitive Psychology, 6, 619-628.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (in press). Verification of face identities from images captured on video. Journal of Experimental Psychology: Applied.

Bruce, V., & Humphreys, G. W. (1994). Recognizing objects and faces. In V. Bruce, & G. W. Humphreys (Eds.), Object and Face Recognition (pp. 141-180). Hove, England: Lawrence Erlbaum Associates.

Bruce, V., & Langton, S., (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. Perception, 23, 803-822.

Bruce, V., Valentine, T., & Baddeley, A.D. (1987). The basis of the 3/4 view advantage in face recognition. Applied Cognitive Psychology, 1, 109-120.

Bruce, V., & Young, A. (1998). In the Eye of the Beholder: The Science of Face Perception. Oxford: Oxford University Press.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. British Journal of Psychology, 77, 305-327.

Bruyer, R. (1983). Le Visage et L'Expression Faciale - Approche Neuropsychologique. Brussels: Pierra Mardaga.

Bruyer, R. (1988). Some masked inconstancies about constancy. Brain and Cognition, 7, 374-376.

Bulthoff, H.H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proceedings of the National Academy of Sciences, 89, 60-64.

Burton, A. M, Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor quality video: evidence from security surveillance. Psychological Science, 10, 243-248.

Burton, A.M. (1994). Learning new faces in an interactive activation and competition model. In V. Bruce and G. Humphreys (Eds.), Object and Face Recognition. Lawrence Earlbaum Associates: Hillsdale, USA.

Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. British Journal of Psychology, 81, 361-380.

Campbell, F.W., & Robson, J.G. (1968). Application of Fourier analysis to the visibility of gratings. Journal of Physiology, 197, 551-566.

Cannon, M.W., Hoffmeister, J.W., & Fullenkamp, S.C. (1997). The spatial frequency bandwidth for aircraft identification (abstract). Investigative Ophthalmology and Visual Science (ARVO Supplement), 38, 2997.

Collin, C.A., Liu, C.H., Chaudhuri, A. (submitted). Face recognition is affected by similarity in spatial frequency to a greater degree than object recognition. Submitted to The Journal of Experimental Psychology: Human Perception & Performance.

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.

Damasio, A.R., Damasio, H., & Van Hoesen, G.W. (1982). Prosopagnosia: Anatomical basis and behavioural mechanisms. Neurology, 32, 331-341.

Damasio, A.R., Tranel, D. & Damasio, H. (1990). Face agnosia and the neural substrates of memory. Annual Review of Neuroscience, 13, 89-109.

Darwin, C. (1872). The Expression of the Emotions in Man and Animals [On-line]. New York: D. Appleton and Company. (available: www.malaspina.com/etext/emotionii.htm)

Davies, G., Ellis, H. D., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. Journal of Applied Psychology, 63, 180-187.

Durgin, F.H., & Proffitt, D.R. (1993). Perceptual response to visual noise and display media. Final Report (NASA Rep. No. NASA-NAG2-814). University of Virginia.

Ellis, H.D. (1986). Introduction to aspects of face processing: Ten questions in need of answers. In H.D. Ellis, M.A. Jeeves, F. Newcombe, & A. Young (Eds.), Aspects of Face Processing (pp. 3-13). Boston: Martinus Nijhoff Publishers.

Ellis, H.D., Jeeves, M.A., Newcombe, F., & Young, A. (Eds.). (1986). Aspects of Face Processing. Boston: Martinus Nijhoff Publishers.

Enns, J.T., & Shore, D.I. (1997). Separate influences of orientation and lighting in the inverted-face effect. Perception & Psychophysics, 59, 23-31.

Farah, M.J. (1996). Is face recognition "special"? Evidence from neuropsychology. Behavioural Brain Research, 76, 181-189.

Farah, M.J., Tanaka, J.W., & Drain, H.M. (1995). What causes the face inversion effect? Journal of Experimental Psychology: Human Perception & Performance, 21, 628-634.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. Perception, 12, 195-201.

Fiser, J., Biederman, I., & Cooper, E. E. (1997). To what extent can matching algorithms based on direct outputs of spatial filters account for human object recognition? Spatial Vision, 10, 237-71.

Fraisse, P., & Elkin, E.H. (1963). Etude genetique de l'influence des modes do presentation sur le seuil de reconnaissance d'objets familiers. L'Annee Psychologique, 63, 1-12.

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

Galper, R. E., & Hochberg, J. (1971). Recognition memory for photographs of faces. American Journal of Psychology, 84, 351-354.

Gauthier, I. (1999). Dissecting face recognition: the role of level of categorization and expertise in object recognition. Talk presented at McGill University, Montreal, Canada, March 19th, 1999.

Gauthier, I., Behrmann, M., & Tarr, M.J. (1999). Can face recognition really be dissociated from object recognition? Journal of Cognitive Neuroscience., 11, 349-370.

Gauthier, I., & Tarr, M.J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. Vision Research, 37, 1673-1682.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., & Gore, J.C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nature Neuroscience, 2, 568-573.

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

Ginsburg, A.P. (1978). Spatial information in a hierarchy of filtered images based upon biologic data (abstract). Investigative Ophthalmology and Visual Science (ARVO Abstract Book), 19, 265.

Ginsburg, A.P. (1980). Specifying relevant spatial information for image evaluation and display design: An explanation of how we see certain objects. Proceedings of the SID, 21, 219-227.

Ginsburg, A.P. (1984). The spatial bandwidth for the identification of geometric form (abstract). Investigative Ophthalmology and Visual Science (ARVO Supplement), 25, 97.

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

Goldstein, A. (1983). Behavioural scientists' fascination with faces. Journal of Nonverbal Behavior, 7, 223-225.

Hancock, P. J. B., Bruce, V., & Burton, M. A. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. Vision Research, 38, 2277-2288.

Harmon, L. D. (1971). The recognition of faces. Scientific American, 229, 70-82.

Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: effects of two-dimensional filtered noise. Science, 180, 1194-1197.

Haxby, J.V., Ungerleider, L.G., Clark, V.P., Schouten, J.L., Hoffman, E.A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. Neuron, 22, 189-199.

Hayes, A. (1988). Identification of two-tone images: Some implications for high- and low-spatial-frequency processes in human vision. Perception, 17, 429-436.

Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. Perception, 15, 595-602.

Hill, H., & Bruce, V. (1993). Independent effects of lighting, orientation, and stereopsis on the hollow-face illusion. Perception, 22, 887-897.

Hill, H., & Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. Journal of Experimental Psychology: Human Perception and Performance, 22, 986-1004.

Hole, G.J., George, P.A., Dunsmore, V. (1999). Evidence for holistic processing of faces viewed as photographic negatives. Perception, 28, 341-359.

Hugdahl, K., Bronnick, K., Kyllingsbaek, S., Law, I., Gade, A., & Paulson, O.B. (1999). Brain activation during dichotic presentations of consonant-vowel and musical instrument stimuli: A -sup-1-sup-50-PET study. Neuropsychologia, 37, 431-444.

Johnston, A., Hill, H., & Carman, N. (1992). Recognizing faces: effects of lighting direction, inversion, and brightness reversal. Perception, 21, 365-375.

Kalocsai, P., & Biederman, I. (1998). Differences of face and object recognition in utilizing early visual information. In H. Wechsler, J. P. Phillips, V. Bruce, F. Fogelman, Soulie, & T. S. Huang (Eds.), Face Recognition: From Theory to Applications. Berlin, Germany: Springer-Verlag.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience, 17, 4302-4311.

Keenan, P. A., Witman, R., & Pepe, J. (1989). Hemispheric asymmetry in the processing of high and low spatial frequencies: A facial recognition task. Brain & Cognition, 11, 229-237.

Keenan, P. A., Witman, R., & Pepe, J. (1990). Hemispheric asymmetry in the processing of high and low spatial frequencies: A facial recognition task erratum. Brain & Cognition, 13, 130.

Kemp, R., McManus, I. C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. Perception, 19, 531-554.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues. Perception, 25, 37-52.

Kemp, R., Towell, N. & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. Applied Cognitive Psychology, 11, 211-222.

Krouse, F.L. (1981). Effects of pose, pose change, and delay on face recognition performance. Journal of Applied Psychology, 66, 651-654.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg., Wurtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions On Computers, 42, 300-311.

Leder, H. (1996). Line drawings of faces reduce configural processing. Perception, 25, 355-366.

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

Liu, C. H., & Chaudhuri, A. (1998). Are there qualitative differences in face recognition between positive and negative? Perception, 27, 1107-1122.

Liu, C.H., Collin, C.A., Rainville, S.J.M., & Chaudhuri, A. (2000). The effects of spatial frequency overlap on face recognition. Journal of Experimental Psychology: Human Perception & Performance, 26, 956-979.

Liu, Z., Kersten, D., & Knill, D.C. (1999). Dissociating stimulus information from internal representation--a case study in object recognition. Vision Research, 39, 603-612.

Logie, R.H., Baddeley, A.D., & Woodhead, M.M. (1987). Face recognition, pose and ecological validity. Applied Cognitive Psychology, 1, 53-69.

Lowe, D.G. (1987). Three-dimensional object recognition from single two-dimensional images. Artificial Intelligence, 31, 355-395.

Luria, S. M., & Strauss, M. S. (1978). Comparison of eye movements over faces in photographic positives and negatives. Perception, 7, 349-358.

Majaj, N., Palomares, M., Mouchraud, C., Kotlyarenko, D. & Peli, D. (1998). Identifying letters--bottom up (abstract). Investigative Ophthalmology and Visual Science (ARVO Supplement), 39, 408.

Marr, D. (1982). Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information. San Francisco: W.H. Freeman.

McCarthy, G., Puce, A., Gore, J.C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus, Journal of Cognitive Neuroscience, 9, 605-610.

Metha, A.B., Vingrys, A.J., & Badcock, D.R. (1993) Calibration of a color monitor for visual psychophysics. Behavior Research Methods, Instruments, & Computers, 25, 371-383.

Millward, R. & O'Toole, A. (1986). Recognition memory transfer between spatial-frequency analyzed faces. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young. (Eds.), Aspects of Face Processing. Boston, MA: Martinus Nijhoff Publishers.

Moore, B.C.J. (1995). Frequency analysis and masking. In B.C.J. Moore (Ed.) Hearing. Handbook of Perception and Cognition (2nd ed.), 161-205.

Morgan, M.J., & Watt, R.J. (1984). Spatial frequency interference effects and interpolation in vernier acuity. Vision Research, 24, 1911-1919.

Morrone, M.C., & Burr, D.C. (1994). Visual capture and transparency in blocked images. Perception, 23, 20.

Morrone, M.C., Burr, D.C., & Ross, J. (1983). Added noise restores recognizability of coarse-quantized images. Nature, 305, 226-228.

Moscovitch, M. (1988). Further analyses of masking functions in laterality studies of face-perception. Brain & Cognition, 7, 377-380.

Moscovitch, M., & Radzins, M. (1987). Backward masking of lateralized faces by noise, pattern, and spatial frequency. Brain & Cognition, 6, 72-90.

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

Norman, J., & Ehrlich, S. (1987). Spatial frequency filtering and target identification. Vision Research, 27, 87-96.

Olds, E.S., Cowan, W.B., & Jolicoeur, P. (1999). Effective color CRT calibration techniques for perception research. Journal of the Optical Society of America, 16, 1501-1505.

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, noise and the efficiency of letter discrimination. Vision Research, 31, 1399-1415.

Parker, D. M., Lishman, J. R., & Hughes, J. (1996). Role of coarse and fine spatial information in face and object processing. Journal of Experimental Psychology: Human Perception & Performance, 22, 1448-1466.

Parr, L.A., Dove, T., & Hopkins, W.D. (1998). Why faces may be special: Evidence of the inversion effect in chimpanzees. Journal of Cognitive Neuroscience, 10, 615-622.

Patterson, K.E., & Baddeley, A.D. (1977). When face recognition fails. Journal of Experimental Psychology: Learning, Memory and Cognition, 3, 406-417.

Peli, E. (1992a). The perception and interpretation of highpass filtered images. Optical Engineering, 31, 74-81.

Peli, E. (1992b). Display nonlinearity in digital image processing for visual communications. Optical Engineering, 31, 2374-2382.

Peli, E., Lee, E., Trempe, C. L., & Buzney, S. (1994). Image enhancement for the visually impaired: the effects of enhancement on face recognition. Journal of the Optical Society of America A-Optics & Image Science, 11, 1929-1939.

Pelli, D.G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. Vision Research, 31, 1337-1350.

Penhune, V.B., Zatorre, R.J., & Evans, A.C. (1998). Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. Journal of Cognitive Neuroscience, 10, 752-766.

Phelps, M.T., & Roberts, W.A. (1994). Memory for pictures of upright and inverted primate faces in humans (Homo sapiens), squirrel monkeys (Saimiri sciureus), and pigeons (Columba livia). Journal of Comparative Psychology, 108(2), 114-125.

Phillips, R. J. (1972). Why are faces hard to recognize in photographic negative? Perception & Psychophysics, 12, 425-426.

Pointon, C. A. (1993). "Gamma" and its disguises: The nonlinear mappings of intensity in perception, CRTs, film, and video. SMPTE Journal, 102, 1099-1108.

Puce, A., Allison, T., Asgari, M., Gore, J.C., & McCarthy G. (1996). Face-sensitive regions in extrastriate cortex studied by functional MRI. Neurophysiology, 74, 1192-1199.

Puce, A., Allison, T., Gore, J.C., & McCarthy, G. (1995). Differential sensitivity of human visual cortex to faces, letter strings, and textures: A functional magnetic resonance imaging study. Neurophysiology, 74, 1192-1199.

Puce, A., Allison, T., Spencer, S.S., Spencer, D.D., & McCarthy, G. (1997). Comparisons of cortical activation evoked by faces by intracranical field potentials and functional MRI: Two case studies. Human Brain Mapping, 5, 298-305.

Pullan, L., & Rhodes, G. (1996). Why are inverted faces hard to recognize? A test of the relational feature hypothesis. New Zealand Journal of Psychology, 25, 8-10.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representation of faces. Cognitive Psychology, 19, 473-497.

Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. Human Neurobiology, 3, 209-222.

Rolls, E. T., Baylis, G. C., & Hasselmo, M. E. (1987). The responses of neurons in the cortex in the superior temporal sulcus of the monkey to band-pass spatial frequency filtered faces. Vision Research, 27, 311-326.

Rolls, E. T., Baylis, G. C., & Leonard, C. M. (1985). Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. Vision Research, 25, 1021-1035.

Rubin, G.S., & Seigel, K. (1984). Recognition of low-pass faces and letters (abstract). Investigative Ophthalmology and Visual Science (ARVO Supplement), 25, 96.

Ryan, T.A., & Schwartz, C.B. (1956). Speed of perception as a function of mode of representation. American Journal of Psychology, 69, 193-199.

Scharf, B. (1971). Fundamentals of auditory masking, Audiology, 10(1), 30-40.

Schuchard, R.A., & Rubin, G.S. (1989). Face identification of band-pass filtered faces by low-vision observers (abstract). Investigative Ophthalmology and Visual Science (ARVO Supplement), 30, 396.

Schwartz, O., Bayer, H., & Pelli, D. (1998). Features, frequencies, and facial expressions (abstract). Investigative Ophthalmology and Visual Science (ARVO Supplement), 39, 173.

Schyns, P.G. & Bulthoff, H.H. (1994). Viewpoint dependence and face recognition. Proceedings of the XVI Meeting of the Cognitive Science Society, 789-793.

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

Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. The Quarterly Journal of Experimental Psychology, 46A, 225-245.

Tanaka, J.W., & Gauthier, I. (1997). Expertise in object and face recognition. The Psychology of Learning and Motivation, 36, 83-125.

Teuber, H.L. (1968). Disorders of memory following penetrating missile wounds of the brain. Neurology, 18, 287-288

Tieger, T., & Ganz, L. (1979). Recognition of faces in the presence of two-dimensional sinusoidal masks. Perception & Psychophysics, 26, 163-167.

Tomonaga, M. (1994). How laboratory-raised Japanese monkeys (Macaca fuscata) perceive rotated photographs of monkeys: Evidence for an inversion effect in face perception. Primates, 35, 155-165.

Tovee, M.J. (1998). Is face processing special?. Neuron, 21(6), 1239-1242.

Tovee, M.J., & Cohen-Tovee, E.M. (1993). The neural substrates of face processing models: A review. Cognitive Neuropsychology, 10, 505-528.

Troje, N.F., & Bulthoff, H.H. (1996). Face recognition under varying poses: The role of texture and shape. Vision Research, 36, 1761-1771.

Troje, N.F., & Bulthoff, H.H. (1998). How is bilateral symmetry of human faces used for recognition in novel views? Vision Research, 36, 79-89.

Tyler, C.W., & McBride, B. (1997). The Morphonome image psychophysics software and a calibrator for Macintosh systems. Spatial Vision, 10, 479-484.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. Cognition, 32, 193-254.

Uttal, W.R., Baruch, T., & Allen, L. (1995a). The effect of combinations of image degradations in a discrimination task. Perception & Psychophysics, 57, 668-681.

Uttal, W.R., Baruch, T., & Allen, L. (1995b). Sequential image degradations in a recognition task. Perception & Psychophysics, 57, 682-691.

Uttal, W.R., Baruch, T., & Allen, L. (1997). A parametric study of face recognition when image degradations are combined. Spatial Vision, 11, 179-204.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. British Journal of Psychology, 79(4), 471-491.

Vermeire, B.A., & Hamilton, C.R. (1998). Inversion effect for faces in split-brain monkeys. Neuropsychologia, 36, 1003-1014.

Wandell, B.A.(1995). Foundations of Vision. Sinauer Associates: Sunderland, Massachusetts.

Watt, R. J. (1988). Visual Processing: Computational, Psychophysical and Cognitive Research. London: Laurence Erlbaum.

Watt, R. J. (1991). Understanding Vision. London: Academic Press.

Weiss, Y., & Edelman, S. (1995). Representation of similarity as a goal of early visual processing. Network, 6, 19-41.

Wiskott, L., Fellous, J-M., Kruger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. Proceeding IEEE International Conference on Image Processing, Santa Barbara, IL, 129-132.

Wogalter, M.S., & Laughery, K.R. (1987). Face recognition: Effects of study to tests maintenance and change of photographic mode and pose. Applied Cognitive Psychology, 1, 241-253.

Wright, A.A., & Roberts, W.A. (1996). Monkey and human face perception: Inversion effects for human faces but not for monkey faces or scenes. Journal of Cognitive Neuroscience, 8, 278-290.

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

Yin, R.K. (1970). Face recognition by brain-injured patients: A dissociable ability? Neuropsychologia, 8, 395-402.

Young, A. (Ed.). (1998). Face and Mind. New York: Oxford University Press.

Young, A.W., & Bruce, V. (1991). Perceptual categories and the computation of grandmother. European Journal of Cognitive Psychology, 3, 5-49.

APPENDIX

The following pages show examples of unfiltered stimuli used in Experiments 6 and 7. Both face and chair stimuli are shown.